CHEMICAL SYNTHESIS AND BIOLOGICAL EVALUATION OF CIRCULAR, BRANCHED AND LARIAT OLIGONUCLEOTIDES

Sandra Carriero

Department of Chemistry
McGill University
Montreal, Canada

August 2003

A Thesis submitted to McGill University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Your file Votre référence
ISBN: 0-612-98221-1
Our file Notre référence
ISBN: 0-612-98221-1

Dedicated to my most loyal fans,
my parents Mina and Mike,
for teaching me the power of perseverance.

ABSTRACT

This thesis highlights novel synthetic routes towards the facile synthesis of lariat DNA and RNA oligonucleotides, and the utilities of branched (bNAs) and circular (*i.e.* dumbbell-shaped) nucleic acids for targeting biologically relevant processes (*i.e.* HIV proliferation, alternative RNA splicing) with potential therapeutic applications. In all cases, synthetic oligonucleotide assembly necessitated only commercially available phosphoramidite building-blocks and standard phosphite-triester chemistry with the exception of the branchpoint adenosine synthon (*i.e.* bis-adenosine phosphoramidite) which was synthesized using published methods.

An innovative synthetic strategy for the synthesis and cyclization of a medium-sized (21-nucleotide) DNA lariat starting from a CPG-tethered, convergently synthesized branched DNA (bDNA) molecule was devised in Chapter 2. This synthetic route exploited the differential cleavage rates of two CPG-oligonucleotide tethers, namely the base-labile hydroquinone-O,O'-diacetate (*Q-linker*) and the more robust succinate (*S-linker*) linkages, as well as phosphitylation of the 5'-oligonucleotide terminus and cyclization under standard phosphoramidite coupling conditions to effect new phosphodiester bond construction. The results clearly indicate a disadvantageous correlation between high branching efficiency on a densely loaded CPG and the production of dendrimeric (*i.e.* hyperbranched) oligonucleotide species rather than effective cyclization. Nonetheless, a small amount of regioisomeric 2'-linked and 3'-linked DNA lariat structures were isolated and their identities confirmed by both spectrometric (MALDI-TOF-MS) and enzymatic methods (BSPDE hydrolysis and 2'-phosphodiesterase activity).

Given the entropic disadvantage of synthesizing medium-sized DNA lariats on solid-support using the method described in Chapter 2, unique intramolecular (*i.e.* DNA dumbbells) and intermolecular template-mediated approaches for lariat cyclization commencing with convergently and divergently synthesized bDNAs and bRNAs were developed in Chapter 3. Both methods lead to the exclusive and high-yielding formation of medium sized (46-57 nucleotides) DNA and RNA lariats. Parameters for successful

phosphodiester bond construction were also elucidated in both systems. The identities of lariat DNAs and RNAs were systematically ascertained using a combination of spectroscopic (*i.e.* T_m and CD analysis), spectrometric (*i.e.* MALDI-TOF-MS) and enzymatic assay (*i.e.* BSPDE and 2'-phosphodiesterase debranching) methods.

A novel class of highly specific and potent oligonucleotide-based HIV-1 reverse transcriptase inhibitors, RNA dumbbells, comprising of a 10 base-pair stem and two flanking UUCG hairpin-loop motifs are described in Chapter 4. Explicitly, such constructs were capable of selectively hampering the RNase-H mediated activity of the retroviral enzyme without consequence to its DNA polymerase function with an IC_{50} in the 3 μM range. Its precise interaction with the RNase H domain of RT was authenticated *via* a UV-crosslinking assay. Furthermore, the RNA dumbbells did not inflict any effect on mammalian RNase H activity, suggesting that such compounds would not obstruct cellular RNase H function.

Chapter 5 describes the utility of synthetic bRNA for the inhibition and modulation of pre-mRNA splicing in yeast and mammalian *in vitro* systems. Most notably, synthetic bNAs can be suitably exploited as agents for the study of branchpoint recognition during *in vitro* splicing of a pre-mRNA transcript. The results clearly indicate the requirement for a fully formed branchpoint (*i.e.* 5'-, 2'- and 3'-extensions; Y-shaped molecules) off the conserved branchpoint adenosine for efficient splicing inhibition. Specific methods for stabilizing bNAs against ubiquitous cellular exo- and endonucleases as well as the 2'-scissle (2'-debranching) activity present in the HeLa extract milieu are also described. Additionally, linear and branched antisense oligonucleotides complementary to the anti-apoptotic Bcl-x_L 5'-splice site and enclosing spliceosomal and ribonucleoprotein-binding overhangs were also studied for their effect on modulating Bcl-x alternative splicing. Many of the branched RNA oligonucleotides represent the largest bNAs ever synthesized by any group. The preliminary results indicate that such oligonucleotides are remarkably effective modulators or Bcl-x alternative splicing by downregulating anti-apoptotic Bcl-x_L splicing and concomitantly upregulating pro-apoptotic Bcl-x_S production.

RESUMÉ

Cette thèse présente une nouvelle approche pour la synthèse facilitée de lassos d'oligonucléotides ADN ou ARN, et démontre l'utilité des acides nucléiques branchés (ANb) et circulaires pour le ciblage de processus biologiques importants (*ex* : la prolifération du VIH, l'épissage alternatif de l'ARN) ainsi que leur potentiel thérapeutique. Toutes les synthèses ont été réalisées à partir de fragments de phosphoramidites commerciaux en utilisant la méthode standard aux phosphotriesters, avec cependant une exception concernant le synthon adénosine du point de branchement (phosphoramidite bis-adénosine) qui a été synthétisé à partir de méthodes publiées.

Une stratégie innovatrice permettant la synthèse convergente d'un ADN branché de taille moyenne (21 nucléotides), attaché à la CPG puis cyclisé en lasso, est décrite dans le chapitre 2. Cette voie de synthèse joue, d'une part, sur les différentes cinétiques de clivage des deux fonctions d'ancrage reliant les oligonucléotides à la CPG, qui sont respectivement l'hydroquinone-O,O'-diacétate (séquence de liaison Q), labile en milieu basique et le succinate (séquence de liaison S) qui est plus stable et, d'autre part, sur la phosphitylation du 5'-terminal de l'oligonucléotide en utilisant des conditions standards de couplage de phophoramidites qui permettent la formation de la liaison phosphodiesters. Les résultats obtenus démontrent clairement qu'une réaction de branchement à haut rendement sur une CPG densément fonctionnalisée, conduit à des oligonucléotides dendrimères (hyper branchés) au lieu d'une cyclisation. De plus, une faible quantité de régioisomères de lassos ADN liés en position 2' et 3' a été isolée et caractérisée par spectrométrie de masse (MALDI-TOF-MS) ainsi que par des méthodes enzymatiques (hydrolyse BSPDE et par l'activité de la 2'-phosphodiesterase).

Étant donné le désavantage entropique à synthétiser des lassos ADN de taille moyenne sur support solide selon la méthode décrite dans le chapitre deux, une approche originale de cyclisation intramoléculaire (*par ex* : dumbbells) et intermoléculaire de brins branchés d'ADN ou ARN synthétisés de manière convergente et divergente est développée dans le chapitre 3. Chacune des méthodes conduit à la formation abondante de lassos ADN et

ARN de taille moyenne (46-57 nucléotides). Les paramètres pour former correctement la liaison phosphodiesters ont été établis pour chacun des systèmes. La caractérisation des lassos ADNs et ARNs a été systématiquement effectuée à l'aide de méthodes spectroscopiques (T_m et analyse CD), spectrométriques (MALDI-TOF-MS) et enzymatiques (BSPDE et le débranchement par l'enzyme 2'-phosphodiestérase).

Une nouvelle classe d'oligonucléotides hautement spécifiques et potentiellement inhibiteurs de la transcriptase reverse HIV-1, les ARN sous forme 'dumbbells', constitués d'un segment de 10 paires de bases et de motifs d'épingles à cheveux UUCG sont décrits dans le chapitre 4. Plus précisément, ces structures sont capables de gêner sélectivement l'activité ARNase de l'enzyme rétro virale avec un IC_{50} de 3 μM, sans perturber les fonctions de l'ADN polymérase. Le domaine d'interaction de l'ARNase du RT a été précisément identifié par un test UV de réticulation. En autre, les altères d'ARN n'ont eu aucun effet sur l'activité de la ARNase H mammalienne , laissant penser que de tels composés n'affecteront pas la ARNase H cellulaire.

Dans le chapitre 5, l'utilité des ARNb synthétiques pour l'inhibition et la modulation de l'épissage in vitro du pré ARNm des espèces mammaliennes et des levures sera décrite. Plus particulièrement, nous verrons que les ANbs peuvent être utilisés comme agents pour l'étude du point du branchement de reconnaissance pendant l'épissage in vitro du pré ARNm transcrit. Ces résultats indiquent clairement que l'adénosine du point de branchement est indispensable pour la formation complète de ce point (*par ex*: extensions 5'-, 2'- et 3', molécules en forme de Y) et pour une inhibition efficace de l'épissage. Des méthodes spécifiques pour protéger les ANbs contre les omniprésentes exo- et endonucléases cellulaires ainsi que leur débranchement en position 2' présente dans l'extrait de la HeLa, sont également décrites dans ce chapitre. De plus, les oligonucléotides antisense linéaires et branchés qui sont complémentaires au site d'épissure 5' du Bcl-x_L antiapoptotique, renfermant du spliceosome et des ribonucléoprotéines qui s'y lient, ont été également étudiés pour leur capacité à moduler l'épissage alternatif du Bcl-x. Plusieurs oligonucléotides d'ARN branchés représentent les plus larges ANbs jamais synthétisés par un groupe. Les premiers résultats indiquent que

de tels oligonucléotides sont d'efficaces et remarquables modulateurs d'épissage alternatif du Bcl-x par leur capacité à réduire celui du Bcl-x_L. Par conséquent, cette reduction favorise la production du pro-apoptotique Bcl-x.

ACKNOWLEDGMENTS

I owe my heartfelt and deepest gratitude to my enthusiastic supervisor, Dr. Masad J. Damha, for taking me under his wing, giving me his trust, and the independence to reach my full scientific potential. More than this, his patience, generosity, kindness and support over the last five years will never be forgotten. I am honored to have worked for such a diligent and methodic scientist.

Warm thanks are warranted to Dr. James D. Friesen and Dorian Anglin (University of Toronto, Banting and Best Department of Medical Research) for extending a gracious invitation to learn numerous Molecular Biology techniques in their laboratory. I am indebted to your goodwill and generosity during my visit.

Many thanks to Dr. Benoit Chabot and Jonathan Villemaire (Université de Sherbooke, Département de Microbiologie et d'Infectologie) for their enriching collaborative work on the Bcl-x project. Furthermore, I am grateful to the RNA splicing expertise provided to me by Oliver Kent and Dr. Andrew MacMillan (University of Alberta). I am most appreciative for the generous donations of HeLa nuclear extract by both Dr. Chabot and Dr. MacMillan throughout the course of my doctoral research.

I am truly obliged to Dr. Antisar Hlil for her expertise and running of numerous MALDI spectra, particularly on a moment's notice. An extra special thanks to Dr. Kyung-Lyum Min for being a model scientist, collaborator, technical expert in molecular biology and good friend during the last stretch of my Ph.D. Thanks to Dr. Katya Viazovkina for her assistance in purifying branched trimers and tetramers as well as synthesizing the ara-A branched oligonucleotide. Your help really kept me on schedule.

Special thanks go out to Fred Kluck and Bill Bastian for their help in designing, building and installing the first P-32 facility in the Chemistry department.

I would especially like to acknowledge my wonderful colleagues in Otto Maass Lab 207, both past and present. You have all influenced my stay tremendously. I owe the most sincere appreciation to Maria Mangos for being such a fantastic supporter, friend and confidante over the last five years. I will never forget our long discussions and problem solving rounds on the brown paper napkins at our neighborhood A.L. Van Houtte. A big thanks to Benedicte Patureau for translating the thesis abstract.

I would like to express my sincere appreciation to the McGill University Chemistry Department and its entire technical and departmental staff for making the department run like a well-oiled machine.

McGill University and Les Fonds Québécois de la Recherche sur la Nature et les Technologies (formely FCAR) are gratefully acknowledged for their generous financial support.

This road to success would have been less traveled had it not been for the remarkable strength and support of the two finest people I have ever known, my parents Mina and Mike Carriero. I share everything I have ever accomplished with you, and thank you for giving to me more than I could ever return. You are my sources of inspiration everyday.

Nonna Nicolina, thank you for giving me a place to stay, a clean driveway, a light on at the door and many hot meals over the last four winters. You are my model of strength.

Special thanks to all of my family and friends who have been constant sources of encouragement, laughter and support over my entire academic career.

I owe my sanity and stability to George Limantzakis for his enduring faith in my abilities, positive attitude and virtuous patience. Thank you for being my voice of reason and backbone.

To all those I have unintentionally left unmentioned, thank you.

TABLE OF CONTENTS

ABBREVIATIONS AND SYMBOLS

A	adenosine
Å	angstrom
A_{260}	UV absorbance measured at 260 nm
Ade	adenine
AIDS	Acquired Immunodeficiency Syndrome
AP	alkaline phosphatase
APS	ammonium persulfate
ATT	6-aza-2-thiothymine
BBP	branchpoint binding protein
bDNA	branched DNA
BIS	*N,N'*-methylene-bisacrylamide
bNA	branched nucleic acid
BPB	bromophenol blue
BPS	branchpoint sequence in the pre-mRNA
bRNA	branched RNA
BSA	bovine serum albumin
BSPDE	bovine spleen phosphodiesterase
Bz	benzoyl
C	cytidine
CD	circular dichroism
Ci	Curie
CIAP	calf intestinal alkaline phosphatase
CNBr	cyanogen bromide
CNIm	N-cyanoimidazole
Cyt	cytosine
DCE	1,2-dichloroethane
DCI	4,5-dicyanoimidazole
ddH_2O	doubly-distilled and deionized water
DEPC	diethyl pyrocarbonate

DIPEA	N,N-diisopropylethylamine
DMAP	4-(dimethylamino)-pyridine
DMSO	dimethyl sulfoxide
DMT	dimethoxytrityl
DMT-Cl	dimethoxytrityl chloride
dN (N=A, G, C, T)	2'-deoxynucleoside
DNA	2'-deoxyribonucleic acid
dsDNA	double-stranded DNA
DTT	dithiothreitol
E. coli	bacterium *Escherichia coli*
EDC	1-(3-dimethylaminopropyl)-3-ethylcarbodiimide
EDTA	ethylene-diamine tetraacetate dihydrate
FAB-MS	Fast Atom Bombardment Mass Spectrometry
G	guanosine
Gua	guanine
h	hours
HATU	O-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyl-uronium hexafluorophosphate
HBTU	O-(1H-Benzotriazol-1-yl)-1,1,3,3- tetramethyl-uronium hexafluorophosphate
HEPES	2-[4-(2-hydroxyethyl)-1-piperazine]ethanesulfonic acid
HIV	Human Immunodeficiency Virus
HPLC	high performance liquid chromatography
Hz	hertz
*i*Bu	*iso*-butyryl
J	coupling constant
λ	wavelength
LCAA-CPG	long-chain alkylamine controlled pore glass
M	molar
m/z	mass to charge ratio

MALDI-TOF-MS	Matrix Assisted Laser Desorption Ionization time-of-Flight Mass Spectrometry
MES	2-(N-morpholino)ethanesulfonic acid
µL	microliters
mL	milliliters
MLV	Murine Leukemia Virus
mol	mole
mRNA	messenger (mature) RNA
NBA	*p*-nitrobenzyl alcohol
nm	nanometers
NMI	N-methylimidazole
NMP	N-methylpyrrolidinone
NMR	nuclear magnetic resonance
nt(s)	nucleotide(s)
NTP	nucleoside triphosphate
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PCR	polymerase chain reaction
PEG	polyethylene glycol
PNK	polynucleotide kinase
ppm	parts per million
pre-mRNA	precursor messenger RNA
Pu	purine
Py	pyrimidine
Q-linker	hydroquinone-O,O'-diacetyl linker
®	registered trademark
r.t.	room temperature
R_f	retardation factor (*i.e.* TLC analysis)
rN (N=A G, C, U)	ribonucleoside
RNA	ribonucleic acid
RNase	ribonuclease

RNase	ribonuclease
RNasin	ribonuclease inhibitor
RNP	ribonucleoprotein
R_t	retention time (HPLC)
RT	reverse trancriptase
RT-PCR	reverse transcriptase polymerase chain reaction
S. cerevisiae	*Saccharomyces cerevisiae*
S. pombe	*Schizosaccharomyces pombe*
SDS	sodium dodecyl sulfate
SEC	Size Exclusion Chromatography (Sephadex G-25®)
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
S-linker	succinyl linker
snoRNA	small nucleolar RNA
snRNP	small nuclear ribonucleoproteins
T	thymidine
Taq	*Thermus aquaticus*
TBAF	tetra-*n*-butylammonium fluoride
TBDMS	*tert*-butyl dimethylsilyl
TBE	Tris/boric acid/EDTA buffer
TCA	trichloroacetic acid
TDT	terminal deoxynucleotidyl transferase
TEAA	triethylammonium acetate
TEMED	*N,N, N',N'*-tetramethylethylenediamine
THF	tetrahydrofuran
Thy	thymine
TLC	thin layer chromatography
T_m	thermal melting temperature
™	Trademark
TREAT-HF	triethylamine trihydrofluoride
Tris	2-amino-2-(hydroxymethyl)-1,3-propanediol

U	units (of enzyme)
U	uridine
Ura	uracil
UV	visible
v/v	volume per volume
w/v	weight per volume
XC	xylene cyanol

LIST OF FIGURES

CHAPTER 1

CHAPTER 2:

CHAPTER 3

LIST OF TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 7

CHAPTER 1: GENERAL INTRODUCTION

1.1. NUCLEIC ACID STRUCTURE AND FUNCTION[1,2]

Fifty years ago this year, the seminal yet modestly-stated 900-word article on the double-helical structure of deoxyribonucleic acid (DNA) by James Watson and Francis Crick (**Figure 1.1**) was published in the journal *Nature*.[3] As simply as that, a new revolution in scientific discovery emerged with the concomitant amalgamation of two key scientific disciplines, namely chemistry and the biological sciences.[4] Their pivotal discovery that DNA adopts an intricate, orderly and defined three-dimensional native state, paved the way for modern molecular biology and genetics.[5] Even earlier than this, it was proposed and later confirmed that DNA is the fundamental molecular storage center of all heritable genetic information.[6] The life surrounding us on Earth and the evolution of species is undeniably based on this simple genetic code transmitted by the DNA molecule and its chemical and biophysical behavior. Although DNA is not the only nucleic acid required by living organisms, it is indeed characterized by two unique and fundamental properties; it can direct its own synthesis (*i.e.* replication) and has the innate ability to mutate (**Figure 1.2**). The DNA polymerase enzymes are the underlying molecules that *replicate*

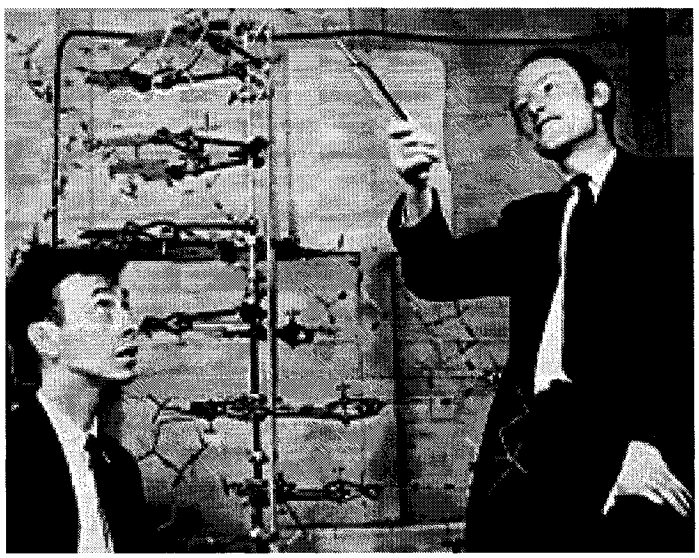


Figure 1.1: Watson (left) and Crick (right) pose with their original DNA model in 1953

DNA through instructions from the parent template in an exquisitely specific fashion. It was later discovered that this transmission of vital genetic information necessitated an intermediary storage center, the ribonucleic acids (RNA)[7], for creation of the true executioner's of genetic function, the protein molecules (**Figure 1.2**). This process of RNA synthesis, the template for protein conception, from a DNA gene is termed *transcription*. The relationship between the coding capabilities of the sequences in an RNA template and the amino acids in proteins, transpires through a process known as *translation*. Other RNA molecules identified as the transfer RNAs (tRNA) and the ribosomal RNAs (rRNAs) are also requisite elements of the protein manufacturing machinery. This orderly relationship from DNA to RNA to protein has been coined the *Central Dogma of Molecular Biology*[8] and is the inevitable pattern by which all genes are expressed in living organisms. Although the genes of most cells and many viruses are comprised of DNA, some viruses (*i.e.* retroviruses) exploit RNA as their genetic material and have a crucial dependence on the host's DNA for viral replication.

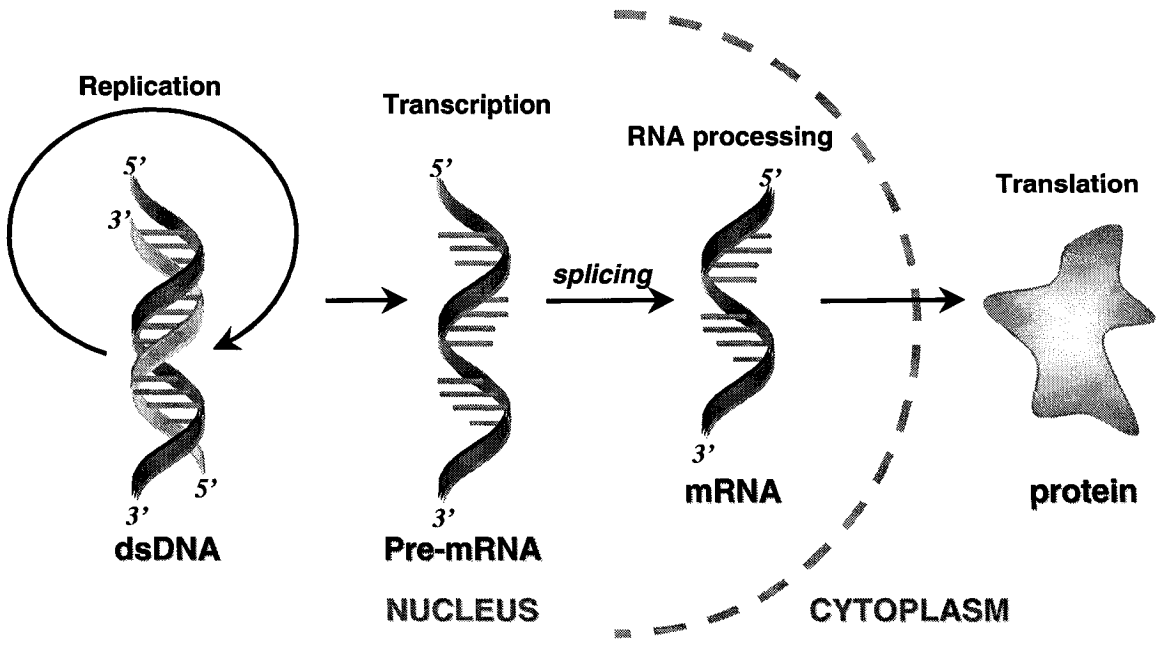


Figure 1.2: Flow of genetic information in eukaryotic cells. DNA replication, transcription of pre-mRNA and RNA splicing all occur in the cell nucleus. The processed RNA is then exported to the cell cytoplasm, where the ribosomal machinery instructs protein synthesis from the mRNA template.

To the unquestioning eye, the DNA double helix appears to be nothing more than two woven strands, much like inter-winding tracks of a roller coaster, held together by innate forces. In reality, the scenario is much more complex than this. Nucleic acids, such as DNA, are long, thread-like polymers, composed of a linear array of monomeric nucleotide subunits. In its native state, the DNA structure comprises of two anti-parallel sugar-phosphate strands, which wind around a central axis in a rigid, right-handed helical arrangement. The strands are held together by a variety of hydrogen-bonds and hydrophobic interactions and the structure is stabilized by base-base stacking interactions among the flanking nucleotide residues at the interior of the helix. Dissection of the individual nucleotide subunits reveals that they comprise of a furanose sugar, a nitrogenous heterocyclic base and a phosphate group linked at the 3'-position of the sugar. The β-D-2-deoxyribose sugar backbone repeats in DNA are linked together *via* 3',5'-phosphodiester linkages, thereby conveying directionality to the strand, and are joined through N-glycosidic linkages with the bases at their 1'-position in a β-configuration (*i.e.* the base lies above the plane of the sugar ring; **Figure 1.3**). Whereas the sugar moieties and phosphate groups operate more of a structural role, it is the sequence of these nitrogenous bases that encodes all the genetic and inherited information in a specific organism, and is contingent upon a four-letter chemical alphabet of A, G, C and T. These planar, heterocyclic bases are divided into two classes; the purines which include adenine (A) and guanine (G) and the pyrimidines, comprising of cytosine (C) and thymine (T). In RNA, the thymine base is replaced by uracil (U) and the sugar in this case is ribose, incorporating a 2'-hydroxyl group, which is not present in the deoxyribose structure. In 1950, Erwin Chargaff came to the conclusion that although the purine and pyrimidine content in diverse organisms varies, the proportion of purines (A and G) to pyrimidines (C and T) is always nearly equal, and signified an element of sequence complementarity within the DNA structure.[9] Indeed, nucleic acids have the innate propensity to recognize each other *via* base-pairing, now known as Watson-Crick base-pairing, through hydrogen-bonding among the heterocyclic bases in opposing strands. The hydrogen-bonding scheme is always represented by pyrimidine-purine base pairs, since such interactions conform to the regular helical nature imposed by the sugar-phosphate backbone in each polynucleotide chain. Specifically, adenine pairs with

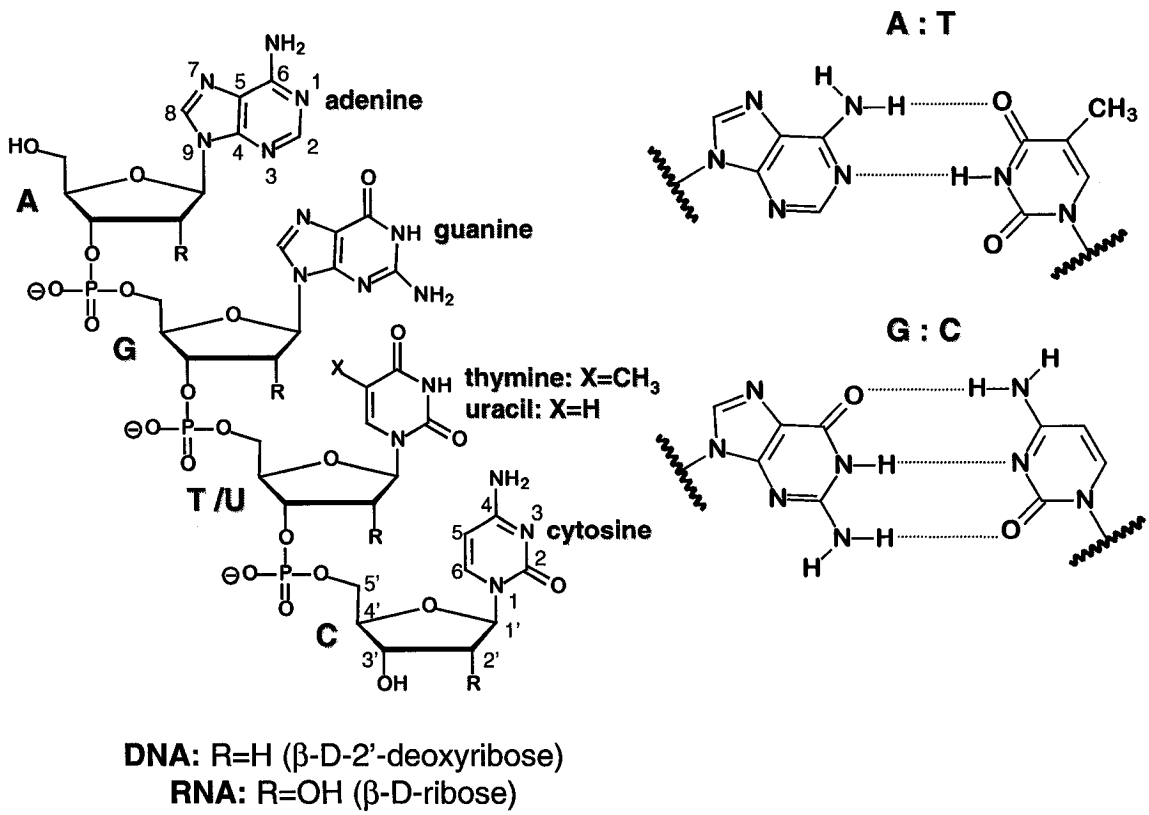
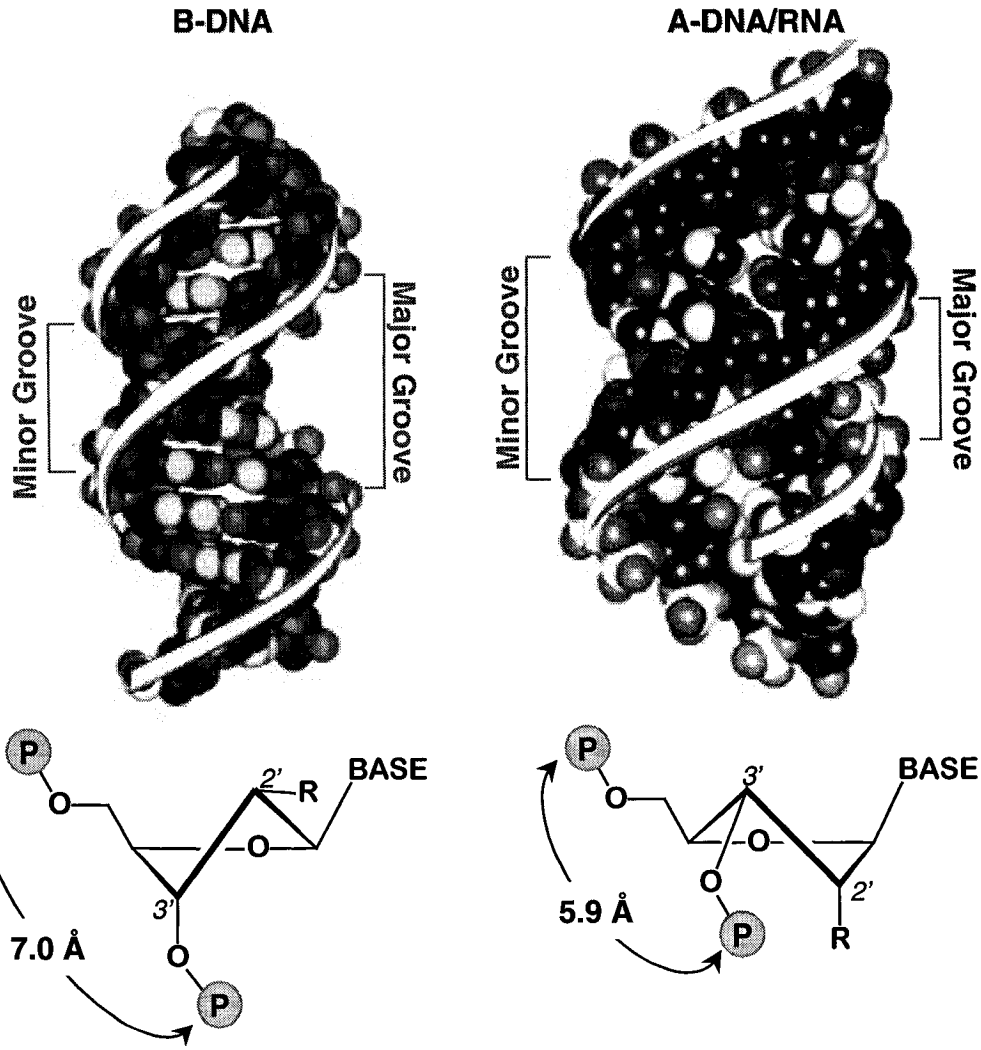


Figure 1.3: Primary structure of 2'-deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) containing the four constituent nucleotides; A, G, T (or U), and C. The conventional numbering used for the carbohydrate carbons, and the pyrimidines and purines is shown (left). To differentiate between the sugar and heterocyclic base positions, the carbon atoms in the furanose ring are primed. Hydrogen-bonding between bases *via* Watson-Crick complementarity (A:T and G:C) is demonstrated on the right.

thymine *via* two hydrogen-bonds, whereas guanine pairs with cytosine *via* three hydrogen-bonds (**Figure 1.3**). This reciprocal recognition of A by T and C by G by way of hydrogen-bonding interactions establishes the reliability of transcription and translation in a given species. A few years ago, studies on DNA polymerase fidelity during replication emerged from Eric Kool's group, and challenged the notion that correct phosphodiester bond formation was contingent upon Watson-Crick base-pairing with the nucleotide in the opposing template strand.[10,11] Indeed, his results demonstrated that nucleotide derivatives which are incompetent at hydrogen-bonding to neighboring bases, yet retain similar sizes and shapes to the natural heterocycles, are capable of being effectively incorporated during DNA strand synthesis by the polymerase.[12,13] By using chemically synthesized DNA precursors, they clearly showed how chemistry can play a

pivotal role in studying vital biological and biochemical processes. Additionally, the chromophoric nature of the bases in DNA as well as RNA are integral in the determination of temperature and pH-dependent changes in base-stacking of the helix as well as in monitoring disruptions in the local asymmetric environment of the heterocycles by a variety of spectrophotometric techniques (*e.g.* ultraviolet spectroscopy (UV), circular dichroism (CD)).[14-16]

The nucleotide units themselves in the helix structure can adopt a variety of different conformations depending on the constituents, and is contingent upon the constraints imposed by the sugar-phosphate backbone and the conformational space accessible to the base-pairs (*i.e.* π-π interactions). As mentioned earlier, DNA in its native state exists as a right-handed double helical structure with the bases pointing inwards, forming hydrogen-bonds and the phosphates and carbohydrate moieties on the exterior, forming a network spine of hydration. The conformational structures of the nucleotides are defined by a series of torsion angles in the phosphate backbone, the furanose ring and in the glycosidic bond.[2] The sugar conformation, or *pucker* as it is usually called, forces the ring bonds to twist out of the pentose plane, thereby minimizing the number of non-bonded interactions between the substituents. This pucker is keenly identified by the displacement of the C2'- and C3'-carbons from the median plane of C1'-O4'-C4' (**Figure 1.4**). In the DNA structure described initially by Watson and Crick, the helix adopts a B-form structure owing to the C2'-*endo* or "S" conformation (C2' is on the same side as C5') adopted by the sugar in a highly humid (low salt) environment. The general attributes of this type of helix are outlined in **Figure 1.4**. Characteristically, the distance between the individual base-pairs along the helix axis is 3.4 Å with a length of approximately 10 base-pairs per turn of the helix (*i.e.* 34 Å long). Additionally, the base-pairs sit directly on the helix axis, thereby producing *major* and *minor grooves* of similar depth. The major and minor groove widths are wide and narrow respectively in the B-form helical arrangement (**Figure 1.4**). The innate flexibility of the deoxyribose sugar moieties allows for the dynamic equilibrium between multiple sugar conformations and is reliant on the helical structure assumed by the duplex. As a result of low humidity (and high salt), DNA can also exist in another polymorphic state known as the A-form. Specifically, this A-DNA



		B-Form	A-Form
Helical sense:		Right	Right
Sugar Conformation:		C2'-*endo*	C3'-*endo*
Helix Diameter:		23.7 Å	25.5 Å
Residues/Turn		10.4	11
Base-Pair displacement from axis center:		-0.2-0.8 Å	4.5 Å
Major Groove	**width:**	11.7 Å	2.7 Å
	depth:	8.8 Å	13.5 Å
Minor Groove	**width:**	5.7 Å	11 Å
	depth:	7.5 Å	2.8 Å

Figure 1.4: Global helical conformations and average structure parameters adopted by DNA and RNA nucleic acids. The preferred sugar conformation of the A-form and B-form structures are depicted below with their internucleotide phosphate-phosphate distances. The averaged structure parameters were adapted from Blackburn, G.M. and Gait, M.J. In *Nucleic Acids in Chemistry and Biology*; Blackburn, G.M. and Gait, M.J., Eds; Oxford University Press: New York, 1996.

polymorph is defined by a shorter internucleotide phosphate-phosphate distance compared to B-DNA owing to the adopted C3'-*endo* sugar pucker (or "N" conformation) in the backbone (**Figure 1.4**). The helix still assumes a global right-handed orientation, however the closer packing of the constituent residues results in 11 base-pairs per turn of the helix, with a vertical rise of 2.6 Å per base-pair. Furthermore, the bases are displaced from the helical axis by 4.5 Å, thereby creating a hollow core down the center of the axis. As a result of these intrinsic features, the major groove in A-DNA is remarkably deep yet narrow whereas the minor groove is extremely shallow but broad. In both the A and B helical arrangements, the glycosidic torsion angle places the bases in an *anti*-conformation, thereby positioning the bulky portions of the heterocycles (*i.e.* the six-membered ring in purines and the O2 in pyrimidines) away from the pentose ring, resulting in a more energetically favored conformation in these structures.[2]

The ribonucleic acids (*i.e.* RNA) on the other hand, exist predominantly in a single-stranded form in the cell. Nonetheless, extensive base-pairing produces co-existant double stranded regions, which are indispensable for the biological function of some RNA species. As aforementioned, RNA bears the same primary structure as DNA (**Figure 1.3**) however the sugar moiety is ribose in this case (2'-hydroxylated) and includes uracil bases instead of thymine. In an identical fashion to thymine, uracil forms base-pairs with adjacent adenine residues, however, it lacks the methyl group at the C5 position of the heterocycle. Furthermore, owing to the presence of the 2'-OH group, RNA duplexes do not exist in B-form helices as this would require a C2'-*endo* sugar pucker, and the result would be steric clashing between the 2'-OH, three atoms of the 3'-phosphate group and the C8-heterocycle atom (when it is a pyrimidine in the *anti*-conformation) in the preceding nucleotide. As such, RNA nucleotides assume a preferred C3'-*endo* conformation promoting the formation of an A-form duplex upon *hybridization* of two complementary RNA strands or regions (**Figure 1.4**). In general, these RNA helices demonstrate similar helical parameters (*e.g.* helical rise, residues per helix turn) to those imparted by the A-DNA duplex. Nonetheless, the extra 2'-OH makes the RNA structure more susceptible to hydrolysis[17], which explains why evolution selected DNA as the true carrier of genetic information

In the heteroduplexed state, the DNA:RNA hybrid assumes neither a pure A-form nor pure B-form helical structure. Instead, an intermediary state between the two forms has been proposed (AB-like or H-form).[18,19] The sugar residues in the RNA strand boast the regular and uniform C3'-*endo* conformation, however those in the DNA strand exist as a mixture of C2'-*endo*, C3'-*endo* and an intermediary O4'-*endo* pucker type.[20,21] This variety of conformations that the deoxyribose can adopt is indicative of its innate flexibility compared to the RNA nucleotides.

1.2. Split Genes and RNA Splicing

In the early days, studies on gene structure and function typically assumed that the only biological significance for RNA molecules was as a passive intermediary storehouse for the messages encoded in DNA, required for effective protein synthesis.[7] In reality, RNA is a very unique biomolecule capable of many more distinct biological functions than its DNA correlative. RNA molecules such as the transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) play integral roles as structural scaffolds during peptide bond formation and protein synthesis. Most notably, these RNA molecules fold into defined structural motifs containing many double-stranded RNA regions.[22] In the case of tRNAs, a high content of unusual nucleosides other than the normal A, G, C and U are present in the molecules such as inosine, pseudouridine, dihydrouridine, ribothymidine and methylated derivatives of guanosine and inosine, and are formed by enzymatic modification of a precursor tRNA. Tom Cech and co-workers initially exposed the catalytic properties of RNAs, with their studies on the self-splicing of a ribosomal RNA precursor in *Tetrahymena thermophila*, a ciliated protozoan, without the assistance of auxiliary proteins.[23,24] Although it was believed for nearly fifty years that all enzymatic activities and chemical transformations in the cell were catalyzed by proteins, Cech's mammoth findings revealed that RNA molecules could catalyze the cleavage and ligation of oligoribonucleotides substrates in a highly specific fashion.[25] Independently, Altman and co-workers demonstrated the true catalytic activity of the RNA moiety in a ribonuclease (RNase P) from both *E. coli* and *B. subtilis*, whereby these RNAs were

capable of effecting the *trans* cleavage of their substrates in the presence of divalent cation (Mg^{2+}) alone.[26-28] Such "ribozymes" were later proposed as a new class of highly specific therapeutic agents, which are capable of cleaving and hence inactivating viral RNAs or other RNAs implicated in disease.[29] Other researchers have further discovered functional roles for RNAs in peptide-bond forming reactions in some lower organisms[30] and as transcription factors in eukaryotic species such as the silkworm.[31] The earlier discovery of an RNA-dependent DNA polymerase activity by an enzyme called *reverse transcriptase* (RT) in RNA tumor viruses, challenged the central scientific dogma which stated that the key to expression of any entity was limited to progression from DNA to RNA to protein.[32,33] Indeed, an enzyme that catalyzed the transformation of an RNA progenitor into DNA was exposed, and lent credibility to the hypothesis, which several principal investigators and Nobel Laureates such as Phillip Sharp (1993), Tom Cech (1989) and Sidney Altman (1989) have proposed, namely, that life on Earth commenced with RNA molecules in the so-called pre-biotic "RNA world".[34,35]

Although RNA biological function is much more vast than initially anticipated, one of its most significant roles remains in the transference of vital genetic information encoded in DNA for reliable protein synthesis (**Figure 1.2**). One of the most remarkable scientific discoveries that emerged from molecular studies on eukaryotic gene expression concerned the fact that functional protein synthesis did not arise from a continuous array of protein-coding sequences in the genomic DNA. Independently, Richard Roberts (Cold Spring Harbor Laboratories) and Phillip Sharp (M.I.T.) discovered that the protein coding regions (*i.e.* exons) in DNA contained interrupted regions (*i.e.* introns; non-coding sequences) nestled within the coding segments, which did not arise in the mature messenger RNA product (mRNA).[36,37] In effect, the average cellular gene consists of approximately eight introns, leading to a primary transcript that is typically four to ten times larger than the final mRNA.[38] These intronic portions, which bear unnecessary elements for protein function, are intricately excised post-transcriptionally in the nucleus from a precursor messenger RNA (pre-mRNA) transcript by a biochemical process, which has been suitably referred to as *RNA splicing*. In essence, the genomic DNA of

most eukaryotic organisms is comprised of these *split genes*, which are precisely ligated to form mRNA molecules with specific protein-coding capabilities. Conserving vital information by accurate splicing is a constraint imposed on genetic systems, as a significant number of human diseases are caused by mutations that interfere with correct splicing of an RNA transcript.[39]

As mentioned previously, an autocatalytic method of splicing rRNA precursors in *Tetrahymena thermophila* was uncovered by Cech and co-workers in 1981.[23] This type of self-splicing, categorized as Group I self-splicing introns, is mediated by a external, nucleophilic guanosine co-factor (*i.e.* guanosine, GMP, GDP or GTP) and was the first experimental proof that an RNA molecule can have highly specific catalytic activity, thereby excising its own superfluous RNA sequences.[40] Messenger RNA precursors in the mitochondria of yeast and fungi, in chloroplasts as well as unicellular organisms such as cyanobacteria and proteobacteria also undergo self-splicing.[41,42] These Group II introns require a specific adenylate residue within the intron for splicing to occur, and a unique topology of intron is excised containing vicinal 2',5'- and 3',5'-phosphodiester linkages which defines the branchpoint of the molecules. As will be discussed in the subsequent sections, the mechanism of Group II intron splicing and nuclear pre-mRNA splicing in eukaryotes are exactly the same and indicate a possibility that these two types of introns are evolutionarily related.[43,44] A further splicing type, *trans*-splicing, performed in parasitic trypanosomes and nematodes (*e.g. C. elegans*) as well as other species, occurs by a distinct mechanism in which two separate, incomplete RNA transcripts are ligated, with the concomitant release of the intron in the form of a forked or branched shape.[45] As *trans*-splicing reactions are exceedingly rare in mammalian cells[46], this reaction has been proposed as novel tool for gene therapy (*i.e.* gene replacement) for the treatment of a variety of genetic diseases.[47,48]

1.3. The Spliceosome: A Macromolecular Machine

In contrast to the Group I and II self-spliced introns, nuclear pre-RNA splicing, which occurs post-transcriptionally in most eukaryotic organisms, is not autocatalytic in nature. Instead, commitment of the nuclear RNAs to the splicing pathway requires an elaborate plethora of *trans*-acting RNA and protein factors; a macromolecular complex that juxtaposes and grasps the reactive sequences for correct excision of the intervening RNA regions and ligation of the coding exons to afford a functional mRNA. The pre-mRNA itself contains a variety of consensus sequences that are phylogenetically conserved among various organisms, and serve as recognition signals for the binding of the diverse splicing factors for the chemical steps of intron-excision/exon-ligation to occur.[49] This evolutionarily conserved splicing mechanism occurs *via* two sequential, ATP-independent transesterification reactions, catalyzed by this macromolecular scaffold known as the *spliceosome*.[50] The splicesome, a 40S (yeast)[50] and 60S (human)[51] entity boasts five uridine-rich small nuclear RNAs (U1, U2, U4, U5 and U6 snRNAs) packaged with associated proteins into ribonucleoprotein complexes (snRNPs), and assemble onto the pre-mRNA substrate in an orderly, step-wise manner. The snRNAs recognize their appropriate sequences on the substrate RNA predominantly through Watson-Crick hydrogen bonding. Non-snRNP protein factors, which associate transiently with the spliceosomal components during splicing, are also present in the splicing milieu of which more than 70-100 have been identified to date by proteomic analysis.[52] The conserved snRNP assembly pathway requires the hydrolysis of ATP as an energy source at many stages. Depending on the eukaryotic species being spliced (*e.g.* yeast or mammalian RNA), the length of the introns varies from very short (*e.g.* 70-nt in fission yeast; *S. pombe*)[53] to many hundreds or thousands of nucleotides in higher-ordered mammalian systems. Nonetheless, consensus sequences at three key intronic sites within the pre-mRNAs appear to be evolutionarily conserved (**Figure 1.5A**). The 5'-end of the intron, succeeding the 5'-exon (*i.e.* the 5'-splice site) in yeast is GUA(A/U)GU compared to GUAAGU in mammalian introns. The 3'-end of the intron, preceding the 3'-exon (*i.e.* 3'-splice site) bears the preserved YAG (Y=pyrimidine) at the intron-exon boundary in both organisms. Of particular importance is recognition of the *branchsite* region of

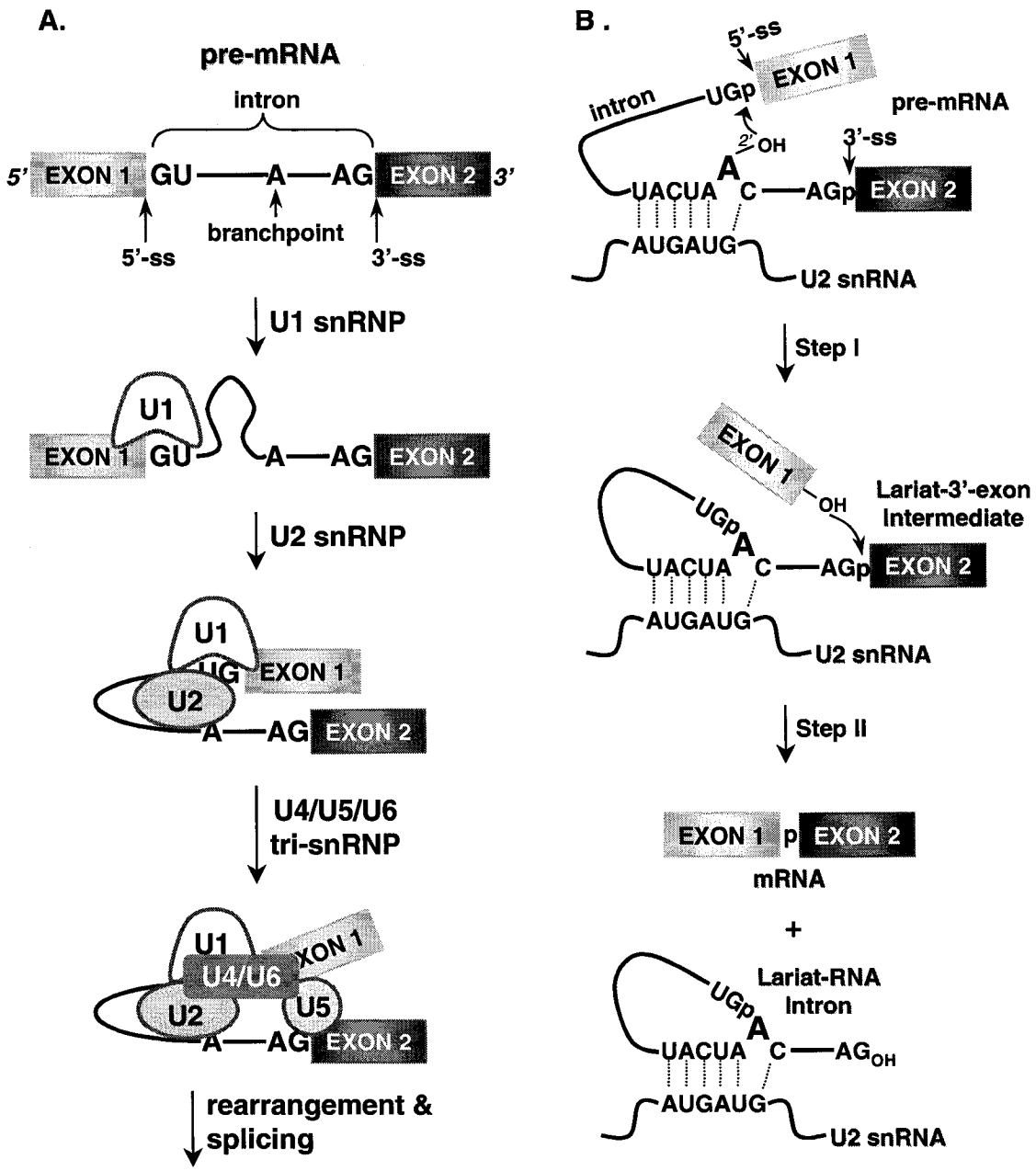


Figure 1.5: Spliceosomal assembly and catalysis. Panel A: Assembly of the individual snRNP particles on a pre-mRNA substrate to form the catalytically active splicing complex. The conserved branchpoint (A), 5'-splice site (GU) and 3'-splice site (AG) are shown. Panel B: Consecutive transesterification reactions catalyzed by the multi-component spliceosome complex. The final outcome results in the ligation of the protein-coding mRNA and release of the branched intron RNA-lariat.

intronic sequences, which is required during early splicing complex assembly and strictly specifies the nucleophile for the first transesterification reaction.[54,55] This branch sequence, UACUAAC (A is the branchpoint nucleophile for the first step of splicing) is stringently conserved in yeast systems, whereas in mammalian systems a more divergent consensus sequence, YURAC is selected (Y=pyrimidines; R=purine), however the yeast sequence remains optimally elected in both systems.[56,57] One key difference in the yeast introns is the lack of a characteristic polypyrimidine tract, located between the branch sequence and the 3'-splice site in mammalian introns.[53,58]

Fundamentally, many of the interactions between the pre-mRNA and spliceosomal components are mutually exclusive, whereby the association of one factor involves the disruption of another, possibly ensuring that the spliceosome can be assembled in an intricately coordinated and carefully regulated fashion. Commitment of the pre-mRNA to the splicing pathways involves initial recognition of the 5'-splice site by the U1 snRNP as well as non-snRNP protein factors such as U2AF (human)[59] or MUD2 (yeast)[60] near the branch region (**Figure 1.5A**). This snRNA-intron interaction was initially proposed in 1980 on the basis of sequence complementarity between the U1 snRNA and the 5'-splice site consensus sequence[61], and later established *in vitro*.[62] Indeed, this interaction has shown to be imperative for commitment of a pre-mRNA to the splicing pathway[63], however, it is not required during the transesterification reactions.[64] Subsequently, the U2 snRNP and its related factors bind to the branchsite region[65] through extensive base-pairing between the U2 snRNA and sequences flanking the branchpoint adenosine (**Figure 1.5A & 1.5B**).[66-68] This early recognition event by the U2 snRNA is proposed to result in the *bulging* of the branchpoint adenosine from U2-intron helix, thereby selecting it as the nucleophile for the first transesterification reaction.[69] The activity of this U2 snRNA appears to be modulated intrinsically *via* a change between disparate U2 conformations.[70] Addition of the associated U4/U5/U6 tri-snRNP is heralded to the rearrangement of numerous RNA-RNA interactions in the spliceosome (**Figure 1.5A**).[71] Specifically, binding of the U5-snRNP to the 5'-splice site in the pre-catalytic complex as well as the 3'-splice site, after the first phosphoryl transfer reaction, is mediated by

snRNP-associated proteins.[72,73] A speculated role for the U1 snRNP in directing the U5 snRNP to the 5'-splice site has been proposed by site-specific crosslinking interactions between the U1 and U5 snRNAs.[74] Most significantly, this triple snRNP is believed to aid in escorting the U6 snRNP to the spliceosome by stable interactions between U4 and U6, possibly by preventing premature activation of U6.[64,75] Indeed, the U6-snRNA is known to make direct contact with both the 5'-splice site and the U2 snRNA (**Figure 1.6**), thereby acting as a molecular scaffold for alignment of the 5'-splice site with the branchpoint in the pre-catalytic complex for the first chemical step of splicing to occur.[76-79]

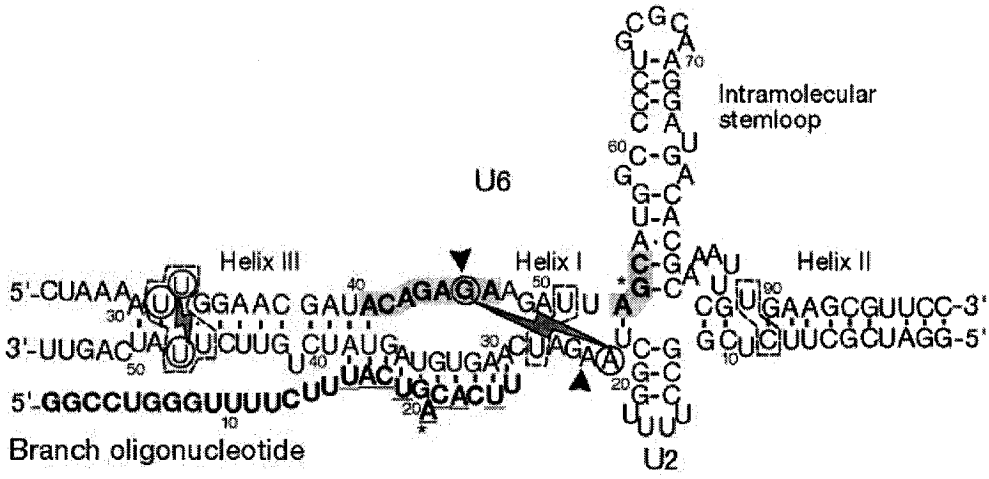


Figure 1.6: Interactions between the pre-mRNA, U2 and U6 snRNAs. A model branch oligonucleotide (Br) is utilized in the place of a full-length pre-mRNA. Shaded boxes mark the invariant regions in U6 and previously established base-paired regions are indicated. The circled residues connected by a zigzag can be crosslinked by UV light. The underlined residues in Br constitute the yeast branch consensus sequence. Asterisks denote the residues involved in the covalent link between Br and U6 snRNA. Adapted from: Valadkhan, S and Manley, J.M. *Nature* **2001**, *413*, 701-707.

Of the five spliceosomal RNAs, the U1 and U4 abandon the spliceosome before the first catalytic step, and the U5 appears to be expendable for at least the first step as well. Specifically, the highly conserved U2 and U6 snRNAs have been implicated in forming the catalytic core for sequential transesterification reactions to occur.[71,77] The likeliness of an RNA catalytic site in the spliceosome is further supported by the mechanistic and

structural similarities between the spliceosome and the autocatalytically splicing Group II introns.[43,80,81] Recently, data emerging from the Manley group has verified the catalytic potential of purified snRNAs in a protein-free system by a UV crosslinking association between invariable sequences in U2 and U6.[82] Additionally, extended work from their lab revealed a remarkable reaction between the 2'-hydroxyl of a bulged adenosine in a branchsite-containing RNA and the catalytically crucial AGC triad in the U6 RNA, which bears many similarities to the first catalytic step of splicing (**Figure 1.6**).[83,84]

Although copious efforts have established many of the significant biological functions of the various spliceosomal and non-spliceosomal components during nuclear pre-mRNA splicing, the overall process remains yet to be defined precisely, especially given the fact that the sequence signals for recruiting the various spliceosomal elements to their appropriate positions on the pre-mRNA substrate are generally weak. Additionally, while the scheme for spliceosomal assembly has been supported by numerous studies in yeast and mammalian cell-based systems, the conditions used for isolation of the various splicing elements (*i.e.* affinity purification) typically do not mirror those that support splicing *in vivo* and *in vitro* (*i.e.* high salt and heparin concentrations).[85-87] Recently, a spectacular report emerging from the laboratory of John Abelson has challenged the stepwise spliceosome assembly dogma, where snRNPs binding to the pre-mRNA in a temporal order.[88] Instead, using gentle affinity purification techniques, they isolated the first penta-associated snRNP complex (U1·U2·U4/U6·U5) complete with 85% of the known factors that demonstrate splicing activity in yeast, suggesting that spliceosomal assembly on a pre-mRNA initiates with a preformed macromolecular complex. In effect, their results indicate that quite possibly the stepwise assembly model is a result of destabilizing experimental conditions which disrupt associations between the native penta-snRNP complex.[89,90] A similar, but less highly characterized preformed entity was also detected many years ago in mammalian splicing extracts.[91] In essence, accurate splice site selections are much more plausible with a preformed complex, especially in more divergent introns (*i.e.* humans), whereby the intron is recognized by an array of

synchronized weak recognition signals. Moreover, this more simplified elucidation of splice site selection will undeniably clarify many of the mechanisms involved for choosing, positioning and matching reactive sites during the splicing reaction.

1.4. NUCLEAR RNA SPLICING: EXCISION OF A LARIAT INTRON

Once the pre-catalytic spliceosomal complex is assembled on the pre-mRNA, splicing of the transcript can occur. The introduction of *in vitro* splicing systems using nuclear extracts and exogenous pre-mRNA transcripts allowed for careful dissection of the events surrounding the splicing reaction.[92] Indeed, *in vitro* studies on the splicing of an adenovirus 2 late major transcript by Sharp and co-workers suggested a likely mechanism for the splicing pathway involving two sequential transesterification reactions.[93] In the first step, the 2'-hydroxyl of the highly conserved branchsite adenosine residue, which presumably bulges out from the pre-mRNA/U2 snRNA duplex[69], performs a nucleophilic attack at the 5'-exon/intron boundary (*i.e.* 5'-splice site), thereby releasing the free 5'-exon and generating a peculiar cyclic intermediate containing the intronic sequences and the 3'-exon (**Figure 1.5B**). This novel type of intermediate has been suitably called a lariat (*i.e.* lariat-3'-exon), owing to its lasso-like topology, and is categorized by vicinal 2',5'- and 3',5'-phosphodiester linkages at the adenosine moiety (*i.e.* branchpoint).[94] The free hydroxyl of the 5'-exon at the 3'-exon/intron periphery in the second splicing step yields the ligated exons (*i.e.* mRNA) with the concomitant release of the lariat-RNA intron, which retains its unique branched architecture (**Figure 1.5B**). As the mechanism is reliant on the two distinct phosphoryl transfer reactions, no exogenous energy source is required. Using phosphorothioate substrates, the stereochemical course of the individual splicing steps has been shown to proceed with inversion of configuration at the phosphorus center, indicating in-line S_N2 nucleophilic displacement reactions.[95,96] Indeed, this is consistent with previous findings obtained for Group I self-splicing introns, and strongly supported the transesterification model for nuclear RNA splicing.[97,98]

Branched RNAs were initially proposed to arise in the cell nucleus during RNA processing by Wallace and Edmonds in 1983[99], however their significance as intronic splicing intermediates and products was not established until one year later in mammalian extract.[94] At approximately the same time, an identical lariat structure was also confirmed in yeast.[100] The lariat species demonstrated a remarkable nuclease resistant structure at the branch core as well as an anomalous mobility when analyzed by denaturing polyacrylamide gels.[101] Moreover, the architecture of the intermediate also imposed a strong block during primer extension analysis, indicating an interruption site at the 3'-end of the intron caused by the branched structure.[94] Characterization of the lariat structure revealed that it was composed of a branched trinucleotide bearing the structure $A^{2',5'}(G)_{3',5'}Y$ (*i.e.* G is linked *via* a 2',5'-phosphodiester bond to the branchpoint A and the pyrimidine (Y) is linked *via* a 3',5'-phosphodiester linkage to A) where the 2'-G is a result of cleavage of the exon/intron boundary at the 5'-splice site, and that the phosphate group at this splice site phosphodiester linkage is incorporated into the lariat structure, thereby releasing the 5'-exon with a free 3'-hydroxyl group (**Figure 1.7**).[102]

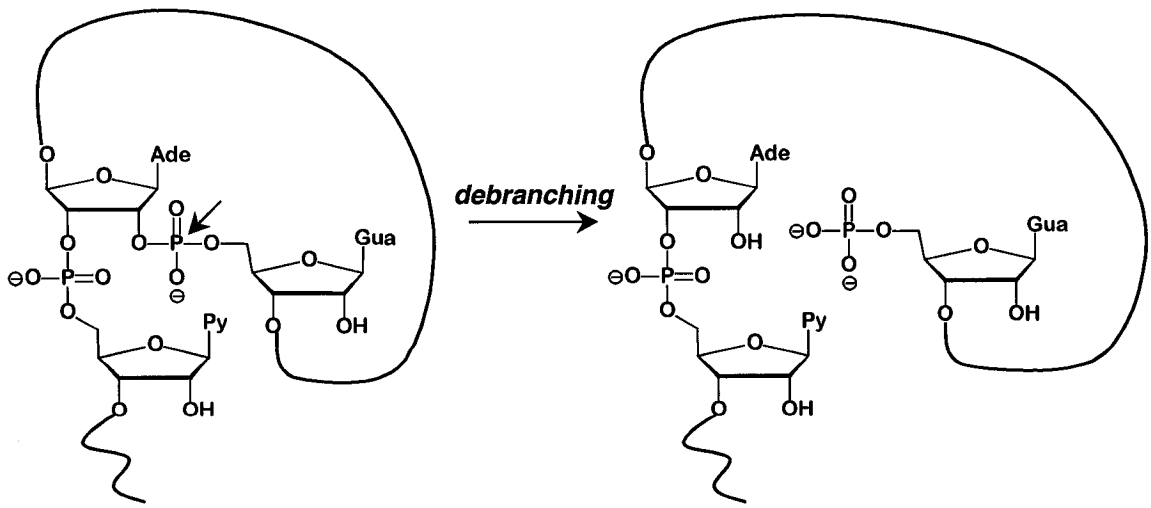


Figure 1.7: Branch core structure of the lariat-RNA intron and 2'-phosphodiester bond metabolism *via* the RNA debranching enzyme. The site of phosphodiester bond hydrolysis is denoted by an arrow (left).

Unequivocally, the selected branchpoint is an invariable adenosine unit, however mutational studies, where the conserved adenosine is replaced for a guanosine unit, revealed that cryptic branchpoint formation at this residues is possible (*i.e* $G^{2',5'}G$).[103] Nonetheless, a significant decrease in the first step of splicing arises (*i.e.* 10-fold decrease in branch formation) as well as a strong block to step two. Additionally, mutation of the normal A branchpoint to G or U, results in the usage of an adenosine residue one nucleotide upstream of the native branchsite, however, a 5-fold reduction in splicing efficiency was also reported.[104] Indeed, these results indicated that the stringently conserved branchpoint adenine base is vital for interactions specific to the second step activation mechanism. Furthermore, the "bulging out" of the adenosine moiety has been implicated as a fundamental element for branchsite selection, through extensive Watson-Crick base-pairing with the U2 snRNA.[66-68] Indeed, site-specific incorporation of 2'-deoxyadenosine or arabinoadenosine units into a pre-mRNA transcript have revealed the significance of both a nucleophilic 2'-hydroxyl and correct sugar-phosphate backbone spatial geometry in the bulged residue, for adequate presentation of the branchpoint nucleophile to the first catalytic splicing step.[69] While cryptic branchpoint selection has been observed in the more divergent mammalian pre-mRNAs[105], single point mutations at the natural adenosine branchpoint in the more strict sequences of the yeast system have been shown to completely abolish splicing altogether.[106]

Previous NMR-derived studies from the Chattopadhyaya group[107], in which the branched core adenosine in a model lariat, a branched tetraribonucleotide, was replaced with cytidine, uridine or guanosine, implicated that divergence of the conformational state at the branchpoint from the canonical (*i.e.* adenosine) may explain why the second step of splicing is severely obstructed in those structures.[103,104] Nonetheless, the small branched structures were not nearly representative of the overall constituent sequences in the native lariats, and as such vital stabilizing interactions by the surrounding residues at the 5'-, 2'- and 3'-extensions off the branchpoint may also be requisite elements for promoting the correct branchpoint conformation to the catalytic portions of the spliceosome. Additionally, modification of the functional groups at the branchpoint

adenine base have revealed a requirement for recognition of this moiety at three steps during the splicing reaction; during early recognition, step one and step two.[108] All three recognitions events appeared to be highly contingent upon N1 of the heterocycle as a hydrogen-bonding acceptor, whereas the presence of a C6-NH_2 group demonstrated the most pronounced activity for early recognition and step one. It remains that the many events that surround branchpoint recognition during splicing need to be investigated in order to paint a clearer picture of why evolution selected the adenosine moiety as the nucleophile for the first chemical step of splicing. Of specific importance is how this native lariat structure is recognized sequentially during the two transesterifications steps, and how this branched topology fits into the overall context of gene expression and genetic diversity.

Analysis of the branched RNA content, later determined to be the intron lariat, in nuclear polyadenylated RNAs, indicated that only 10% of the total RNA material was present in this branched form.[99] This suggested that either the branched molecules were being rapidly degraded by host nucleases, or that specific hydrolysis at the branched core was arising, resulting in the production of linear RNA species. As the branched structures were shown to be highly nuclease resistant[101], it appeared that an alternate nuclease activity, specific to the branched core architecture, may have been enlisted instead. Indeed, in 1985, Ruskin and Green discovered a specific 2',5'-phosphodiesterase activity in HeLa cells, which selectively hydrolyzed the 2',5'-linkage adjacent to a 3',5'-phosphodiester bond, in splicing derived lariat species, thereby linearizing the intron (**Figure 1.7**).[109] This novel activity, which they suitably called the *RNA debranching activity*, was distinct from previously reported 2',5'-phosphodiesterases (*e.g.* snake venom phosphodiesterase) since it necessitated stringent vicinal 2',5'- and 3',5'-phosphodiester linkages for activation. Interestingly, this debranching activity was only observed when the *in vitro* synthesized lariat was completely deproteinized and then added back to a splicing reaction, indicating that the 2',5'-phosphodiester bond in the lariat is protected from degradation, possibly by one or more nuclear factors present in the spliceosome.[65,110,111] This novel debranching activity was later purified 700-fold from the cytosolic fraction of HeLa cells and demonstrated a strict requirement for divalent

cations.[112] Furthermore, branched trinucleotide structures (*i.e.* $pA^{2',5'}(pX)_{3',5'}pY$) prepared by the digestion of RNA lariats with nuclease P1 were also substrates for the debranching activity, however the 2'-phosphate monoester species (*e.g.* $pG^{2',5'}(p)_{3',5'}pC$) was not selectively hydrolyzed. This revealed a remarkable requisite for an intact phosphodiester linkage at the 2'-position in the branched structures. A few years later, the gene encoding for the yeast (*Saccharomyces cerevisiae*) lariat debranching enzyme, *dbr1* was isolated and characterized by Chapman and Boeke.[113] Indeed, the expressed protein from *S. cerevisiae* demonstrated 40% sequence homology to the human debranching enzyme, however it did not require divalent cations for enzymatic activity. Moreover, they showed that eradication of debranching activity *in vivo* resulted in accumulated levels of "nibbled" branched lariat introns (*i.e.* lacking portions of their 3'-tails), suggesting that debranching of the lariat species was a rate limiting step in the intron metabolic pathway. The enzyme was also capable of digesting a variety of branched nucleic acid substrates such as Group II intron lariats, multicopy-single-stranded DNAs (msDNAs)[114] and synthetic branched oligonucleotides.[115,116] A simple hydrogen[109] or hydroxyl group[117] at the 3'-position of synthetic branched structures did not support cleavage, while a phosphate group at the 3'-position sustained hydrolysis extremely inefficiently.[117] Increased levels of lariat introns in *S. cerevisiae* were shown to be non-detrimental to cell viability, however, it was expected that lariat intron accumulation in higher ordered eukaryotes (*e.g.* humans), which contain a larger ratio of intronic to exonic sequences, would be more deleterious to the cell. Assuredly, efficient turnover of excised introns appeared to be much more essential in a higher yeast species (*Schizosaccharomyces pombe*), which contains 40-times more introns than *S.cerevisiae*. An *S. pombe* mutant, deficient in debranching activity, led to severe growth defects in the cell, and an aberrant elongated phenotype.[118] These results implied that either the load of unprocessed introns were physiologically deleterious to the cell, possibly through titration of essential RNA splicing factors specific to the branched architecture. Otherwise, the debranching activity may be responsible for production of linearized introns with specific coding capabilities for other cellular functions. Indeed, an eminent dependence on the RNA lariat debranching activity has been demonstrated in the

biosynthesis of intron-encoded small nucleolar RNAs (snoRNAs) which are directly involved in events that process precursor ribosomal RNAs, (rRNAs) such as pseudouridylation and 2'-O-methylation of rRNA.[119] In mutant *dbr1* cells, immature forms of intronic snoRNAs accumulated, and were trapped within nondebranched intron lariats, indicating a critical dependence on RNA splicing for biogenesis of intronic snoRNAs.[120] Curiously, the *dbr1* mutation did not block rRNA methylation by a U24 snoRNA, suggesting that at least this type of snoRNA is still functional in the form of a branched lariat.

1.5. ALTERNATIVE RNA SPLICING

The discovery that the human genome is comprised of only 35000 genes[121,122] led many to wonder how it is possible that so few genes encode for such immense proteomic diversity (*i.e.* 250000 proteins). The answer, is *alternative RNA splicing*. The simplistic picture painted in the preceding sections of constitutive RNA splicing demonstrates how transcription of one gene results in the production of a pre-mRNA containing only two exons with the concominant extrusion of one intron, consequently allowing for the expression of a single protein product. In reality, up to 60% of human genes[123] as well as the genes of many other species such as *Caenorhabditis elegans* (*i.e.* a nematode) and *Drosophila melanogaster* (*i.e.* the fruit fly) are alternatively spliced such that a gene containing multiple introns and exons can be spliced in a variety of fashions, depending on the splice sites selected, leading to the production of multiple and distinct mRNAs that encode specific protein products (**Figure 1.8**).[124,125] Approximately 80% of alternative splicing results in alterations in the encoded protein as a result of the mRNA variability, which can insert or remove amino acids, shift a reading frame or introduce a premature termination codon.[123] This splicing pathway is often tightly regulated to a specific cell type, with an explicit developmental stage in the cell, or in response to an external stimulus, and is a powerful and versatile regulatory mechanism that can effect quantitative control of gene expression and functional diversity in protein products.

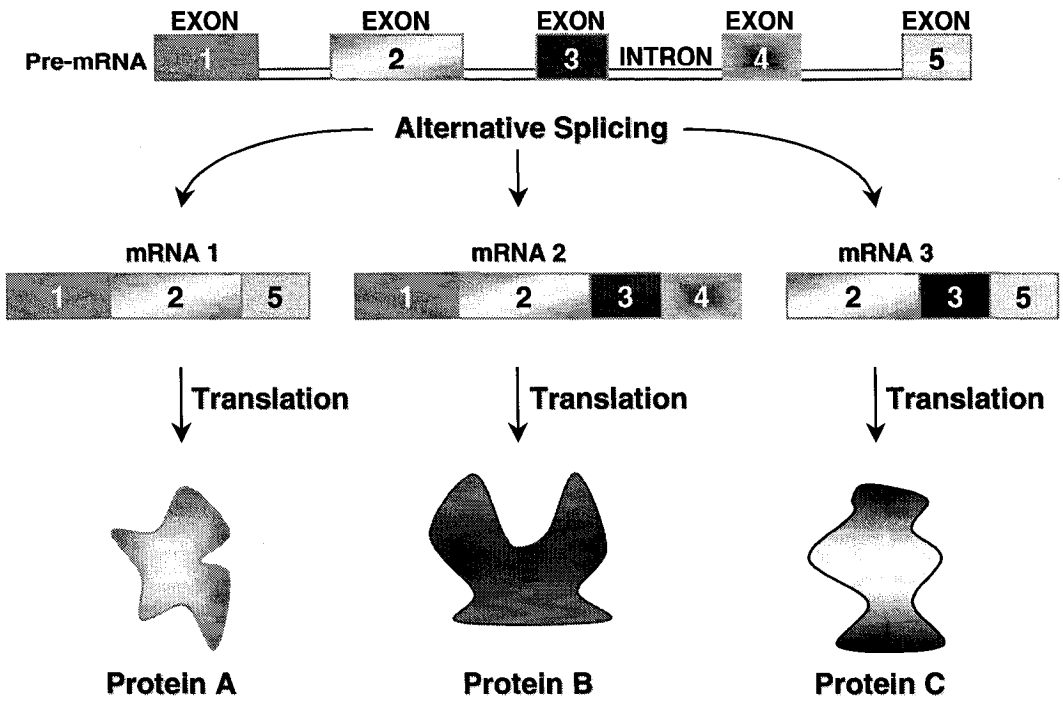


Figure 1.8: Schematic representation of alternative RNA splicing. Alternative splicing of a pre-mRNA transcript results in the generation of multiple and distinct mRNA products which code for diverse proteins from a single gene.

Specifically this splicing program is an integral component of gene expression involved in neuronal differentiation (*e.g.* ion channels, receptors, neurotransmitters)[126], sex determination[127] and apoptosis (*i.e.* programmed cell death).[128]

Splicing by the alternative pathway involves selection of splice sites by the snRNPs, as detailed in the preceding sections, however the mechanisms of selecting competing splice sites in a single gene for alternative splicing to occur are not completely understood yet. The considerable body of evidence indicates that selection of splice sites is controlled by competition between splice sites and/or related splicing sequence elements for factors during the spliceosomal assembly process.[129,130] The mechanism of exon ligation and intron exclusion in both constitutive and alternative splicing are identical in all respects, occurring *via* two sequential transesterification reactions at specific intron/exon junctions. A superior class of extensively phosphorylated splicing factors known as the SR (*i.e.*

serine-arginine rich) family of polypeptides[131] impose an additional regulatory element to alternative splicing by supporting exon definition.[132,133] By far, the most important sequence elements of a pre-mRNA that affect splice site selection are exonic splicing enhancers (*i.e.* promote recognition of a specific splice site)[134], secondary structures in the pre-mRNA itself, and most importantly, matching of the 3' and 5'-splice sites and branchpoints to their respective consensus sequences for adequate recognition. This is well established by the fact that mutations in the pre-mRNA at requisite splice sites can lead to devastating genetic diseases owing to the activation of cryptic sites, through the skipping of necessary exons adjacent to the mutated splice sites, or the inclusion of otherwise disregarded exons.[39,135] In the absence of mutations, other splicing scenarios are possible, potentially leading to the production of proteins with synergistic or even antagonistic functions. For example, many alternatively spliced genes directly related to controlling cellular apoptosis, display antagonistic behaviors. Some of these apoptosis-related genes can be alternatively spliced to include; the full inclusion or exclusion of certain exons (**Figure 1.8**) as in Caspase 2 (Ich-1)[128], selection of an alternative 5'-splice site as with Bcl-x[136], or a 3'-splice site as with the BH3 protein BAX.[137] In the cases described, these alternative splicing pathways can lead to the production of specific mRNAs that encode for proteins with antagonistic functions, such as the anti-apoptotic Bcl-x_L (*i.e.* long isoform; uses an a downstream 5'-splice site in exon 2) and pro-apoptotic Bcl-x_S (*i.e.* short isoform; uses an upstream 5'-splice site in exon 2).[136] The implications of this alternative splicing pathway are highly significant in relation to cells undergoing uncontrolled proliferation (*i.e.* cancer cells) since higher expression of an anti-apoptotic protein (*e.g.* Bcl-x_L) would promote cell survival, whereas increased expression of a pro-apoptotic protein (*e.g.* Bcl-x_S) would counteract this cell endurance thereby stimulating necessary cell death.[138] As such, numerous efforts are currently underway to regulate splice site selection in various apoptosis related genes so as to increase expression of pro-apoptotic proteins, thereby downregulating anti-apoptotic gene expression.[139-143]

1.6. HISTORICAL DEVELOPMENTS IN NUCLEIC ACID CHEMISTRY: VERSATILITY OF THE SOLID-PHASE PHOSPHITE TRIESTER APPROACH

In the early 1950's, even before the double helix structure of DNA was completely elucidated, H. Gobind Khorana, a trained organic chemist, was already looking at versatile methods for synthesizing dinucleotides and oligonucleotides. In 1953, he and Sir Alexander Todd pioneered the utility of the first carbodiimide-based coupling reagents for the preparation of phosphorylated compounds.[144] An extension of this work towards the synthesis of internucleotide linkages using similar carbodiimide-mediated condensation conditions, paved the way for the solution-based *phosphodiester* approach to oligonucleotide synthesis.[145] This method proved indispensable for the preparation of short DNA trinucleotide repeats, which were exploited in deciphering the genetic code, a remarkable achievement for which Khorana shared the Nobel Prize in Physiology and Medicine with colleagues Robert W. Holley and Marshall W. Nirenberg.[146] In 1970, when Khorana and his co-workers revealed their work on the total synthesis of the yeast alanine tRNA through a alliance of chemical and enzymatic methods, the general consensus among scientists was that such synthetic structures would be of little utility to the scientific community.[147] Contrarily, in the words of Tom Cech, 1989 Nobel Laureate for his role in the discovery of the catalytic properties of RNA, "Chemical synthesis of DNA is the fundamental technology that has led to the molecular biology revolution".[4] Indeed, such oligonucleotide structures have proven cardinal as probes and primers in DNA sequencing and genetic engineering, and essential tools for gene therapy approaches and in amplification of genetic material by the polymerase chain reaction (PCR).

Although Khorana's work in oligonucleotide synthesis using the phosphodiester approach was monumental and extremely original for its time, the approach suffered from numerous drawbacks especially tedious and inefficient purification of the charged structures, extensive pyrophosphate side-product formation and chain-cleavage from competing reaction between the internucleotide phosphodiester and activated nucleoside monophosphate, as well as failure to produce sufficiently long oligonucleotides. The

effective limit of this method, an oligonucleotide of 10-15 residues, would take upward of three months to prepare. Significant contributions to the field of oligonucleotide syntheses were further realized by Letsinger's group (Northwestern University) in the late 1960's, to overcome the inevitable disadvantages of Khorana's phosphodiester approach, as well as to adapt a method of synthesizing these biomolecules while tethered to an insoluble polymer support.[148-151] Letsinger's *phosphotriester* approach revolutionized the step-wise assembly of oligonucleotide chains both in solution and on solid-support, affording fully protected and neutral species that were much more effectively purified. Additionally, the introduction of phosphate protecting groups (*i.e.* phosphotriester) such as the chlorophenyl, methyl or β-cyanoethyl groups, eliminated the competing coupling with internucleotide phosphodiesters seen previously.[150,152] Limitations with this method are still palpable owing to the coupling conditions utilized and irreversible phosphorylation reactions, particularly at the O6-position (lactam function) of guanine bases. Nonetheless, this approach still attains widespread use for the large-scale, solution-based synthesis of oligonucleotides.[153]

Milestone contributions by Letsinger and co-workers in the 1970's facilitated internucleotide bond formation through the development of *phosphite triester* chemistry.[154,155] The approach exploited the efficient coupling of highly reactive chlorophosphites with the 5'-hydroxyl of a nucleoside, followed by rapid *in situ* oxidation of the phosphite triesters to the more stable phosphate triesters, and was successfully adapted to the synthesis of both oligodeoxyribonucleotides[156] and oligoribonucleotides.[157-159] Nonetheless, the moisture sensitive chlorophosphite intermediates were very difficult to handle and limited their applicability for automated solid-phase synthesis. A strategic and monumental development to the field of nucleic acid chemistry was the introduction of the nucleoside N,N-dialkyl phosphoramidite synthons by Beaucage and Caruthers (University of Colorado, Boulder) which, when activated by weak acids such as tetrazole (pKa≈4.9) permitted the near quantitative internucleotide coupling with a suitably protected nucleoside in less than one minute.[160,161] Such oligonucleotide building blocks were isolable as stable powders,

which could be readily purified and stored for long time-periods, and were not sensitive to oxidation or hydrolysis. It was this significant advancement in the field that permitted the rapid, automated solid-phase synthesis of oligonucleotides, which had previously been introduced by Ogilvie and co-workers (McGill University) during the early 1980's[162], and later altered to exploit the effective nucleoside phosphoramidite building blocks for efficient chain assembly and commercialization of the "gene machine" for both DNA and RNA synthesis. This phosphoramidite procedure remains the method of choice to this day for the small-scale (microgram to milligram) synthesis of oligonucleotide up to 150 nucleotides long.

A straightforward approach to solid-phase biopolymer synthesis was a revelation initially created by Merrifield[163] and Letsinger[164] for the synthesis of peptides and later, oligonucleotides.[149] The essence of solid-phase synthesis is the heterogeneous coupling reaction between a solublized nucleotide derivative and another nucleoside residue bound to an insoluble resin. The key advantage of this methodology is that the reaction can be forced to high yield by the treatment with excess nucleotide derivative in solution. The reactions take place in one simple vessel, and the excess nucleotide can be simply removed by washing of the resin surface and filtration. This diminishes loss of material due to manual manipulation. With appropriate orthogonal deprotection at various stages, the reactions can be reproducibly carried out in a cycle by flowing solvents and reagents in a predetermined order to assemble polymers of desired length and sequence composition. Automated oligonucleotide synthesizing instruments utilize phosphoramidite chemistry and coupling conditions to assemble sequences on a controlled pore glass (CPG) resin. CPG is much preferred over previously utilized silica gel supports as they are uniformly-sized, rigid particles which retain their pore dimensions after exposure to reagents and do not cause the back-pressure problems inherent to the silica gel matrix. Typically, oligonucleotide sequence assembly begins with a nucleoside loaded onto a solid-support, which becomes the 3'-terminus of the molecules after cleavage of the covalently attached linker arm. The nucleoside is tethered to the support *via* long chain alkylamine (LCAA) spacer arm and base labile linker such as the classically employed dicarboxylic acid, succinic acid[165,166] (**Figure 1.9**) or most

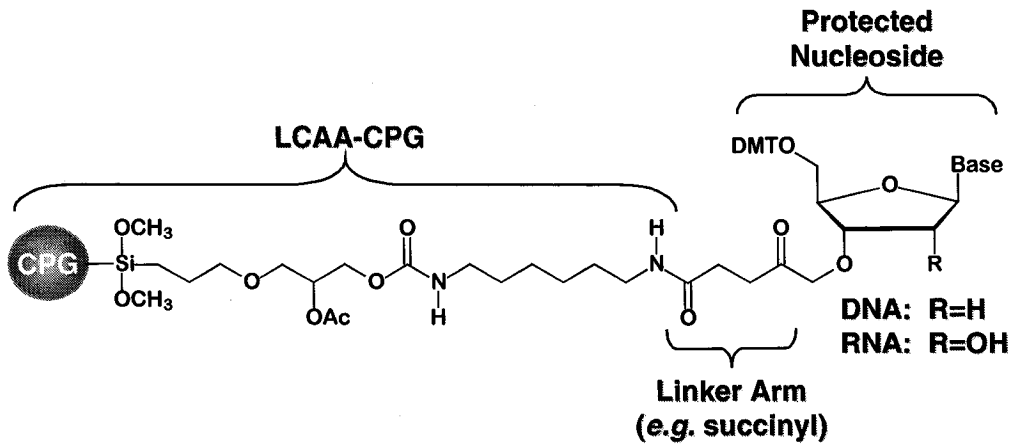


Figure 1.9: Structure of long-chain alkylamine controlled-pore glass (LCAA-CPG) derivatized with a linker arm and suitably protected 3'-nucleoside.

recently, the rapidly cleaved hydroquinone-O,O'-diacetic acid linker.[167,168] Such linkers are stable to the conditions employed for orthogonal deprotection during strand elongation, and readily cleaved by ammonolysis after completion of oligonucleotide synthesis.

The phosphoramidite building blocks suitably employed for both DNA and RNA oligonucleotide synthesis contain distinct protecting groups, which abolish side reactions at either the sugar hydroxyl positions, phosphate and exocyclic amino groups of the heterocyclic bases (**Figure 1.10**). Most of the protecting groups remain covalently attached to their appropriate functionalities until complete deprotection of the oligonucleotide strand at the end of a synthesis. Alternatively, the transient 5'-hydroxyl protecting groups, the 4'-monomethoxytrityl (MMT) or the more labile 4,4'-dimethoxytrityl (DMT) group, are readily eliminated by mild acidic treatment with every successive phosphoramidite coupling cycle to expose the reactive 5'-terminus.[169,170] Heterocyclic base protecting groups such as the benzoyl (Bz) protecting group on adenine and cytosine and the isobutyryl (i-Bu) group on guanine, are introduced by acylation of the exocyclic amino positions, and remain tethered until complete deprotection (**Figure 1.10**).[171] Additionally, the phosphate β-cyanoethyl protecting group is stable towards all the synthesis conditions and remains intact until oligonucleotide deprotection. Existence of an additional 2'-hydroxyl functionality in RNA oligonucleotides necessitates

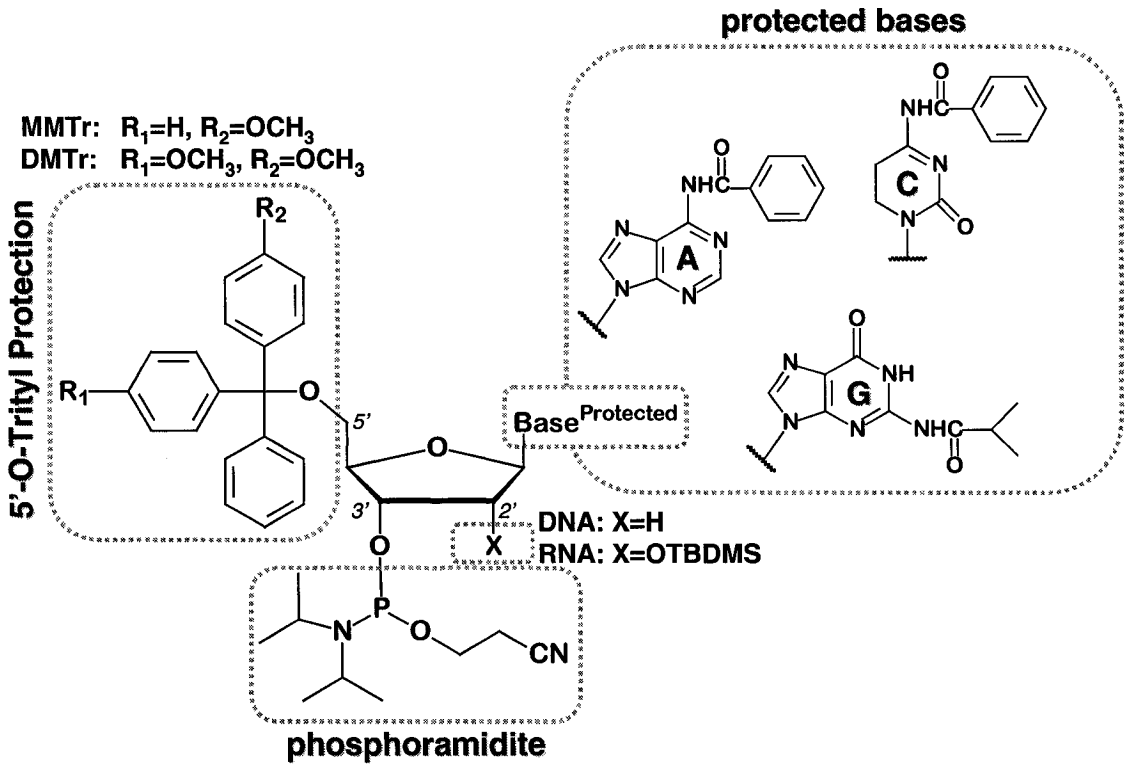


Figure 1.10: Phosphoramidite structure and protecting groups utilized in automated DNA and RNA solid-phase synthesis. DMTr=dimethoxytrityl; MMTr=monomethoxytrityl; TBDMS=*tert*-butyldimethylsilyl. The exocyclic amino protecting group on adenine and cytosine is benzoyl (Bz) while that on guanine is the iso-butyryl (i-Bu) group.

a supplementary protecting group. Principally, this 2'-hydroxyl is protected in the nucleoside phosphoramidites by way of the *tert*-butyl dimethylsilyl (TBDMS) protecting group, an acid and base stable moiety introduced by Ogilvie and co-workers for the effective synthesis of oligoribonucleotides (**Figure 1.10**).[172,173] The TBDMS group, originally described by Corey for the protection of hydroxyl units in prostanglandins, is easily removed under neutral conditions in the presence of fluoride ions.[174] Other 2'-OH protecting groups, such as the acid-labile orthoesters (2'-ACE chemistry)[175] are currently being exploited for research and commercial purposes, yet TBDMS remains the protecting group of choice for routine RNA synthesis.

The conventional automated solid-phase synthesis cycle employed for the assembly of both DNA and RNA oligonucleotide structures *via* phosphoramidite synthons and

phosphite triester chemistry is illustrated in **Figure 1.11**. Cycle initiation commences with the treatment of the nucleoside-loaded CPG with a trichloroacetic acid solution (3% TCA in dichloroethane) in order to release the 5'-protecting group, in this case DMT. This *detritylation* reaction results in the collection of the bright orange trityl cation species (DMT^+), which can be quantified by spectrophotometry to determine the efficiency of successive coupling reactions. Subsequently, the liberated 5'-OH of the tethered nucleoside undergoes a coupling reaction with a DNA or RNA nucleoside 3'-phosphoramidite synthon in the presence of the mildly acidic activating reagent, 1-H-tetrazole, thereby increasing the length of the chain by one unit.[176,177] This coupling results in the production of an intermediary phosphite triester linkage. Although the phosphoramidite approach is renowned for its high coupling efficiency, oligonucleotide chain extension does not occur quantitatively even under the most optimal conditions (98-99% average coupling/addition). As such, the 5'-termini of any unreacted support-tethered nucleotides or oligonucleotides are *capped* by acetylation, thereby preventing the extensive contamination of the final product with a population of shorter (n-1) oligomers, and aiding in product purification.[178] The most commonly used capping solution, a mixture of acetic anhydride, 2,6-lutidine or 2,4,6-collidine, and N-methylimidazole[179] also minimizes the amount of O6-phosphitylated guanine species (*i.e.* tri-coordinated enol phosphites), which arise during the coupling step, and as such, is generally conducted prior to oxidation.[180,181] Oxidation of the phosphite triester intermediate to a more stable pentavalent phosphate triester is accomplished with an aqueous solution of iodine and pyridine in tetrahydrofuran.[154,155] Repetition of the cycle N-times results in the assembly of an oligonucleotide N+1 units long (**Figure 1.11**). The 5'-trityl on the protected product oligonucleotide can either be removed prior to chain cleavage and deprotection, for subsequent purification by anion-exchange HPLC or polyacrylamide gel electrophoresis (PAGE), or retained as a hydrophobic handle for purification by reverse-phase HPLC. Cleavage of the oligonucleotide from the CPG-surface and elimination of the phosphate and exocyclic amino protecting groups under basic conditions (*i.e.* aqueous ammonia) results in the desired, fully-deprotected oligodeoxynucleotide sequence. An ensuing treatment with fluoride reagent such as *tert*-butylammonium fluoride

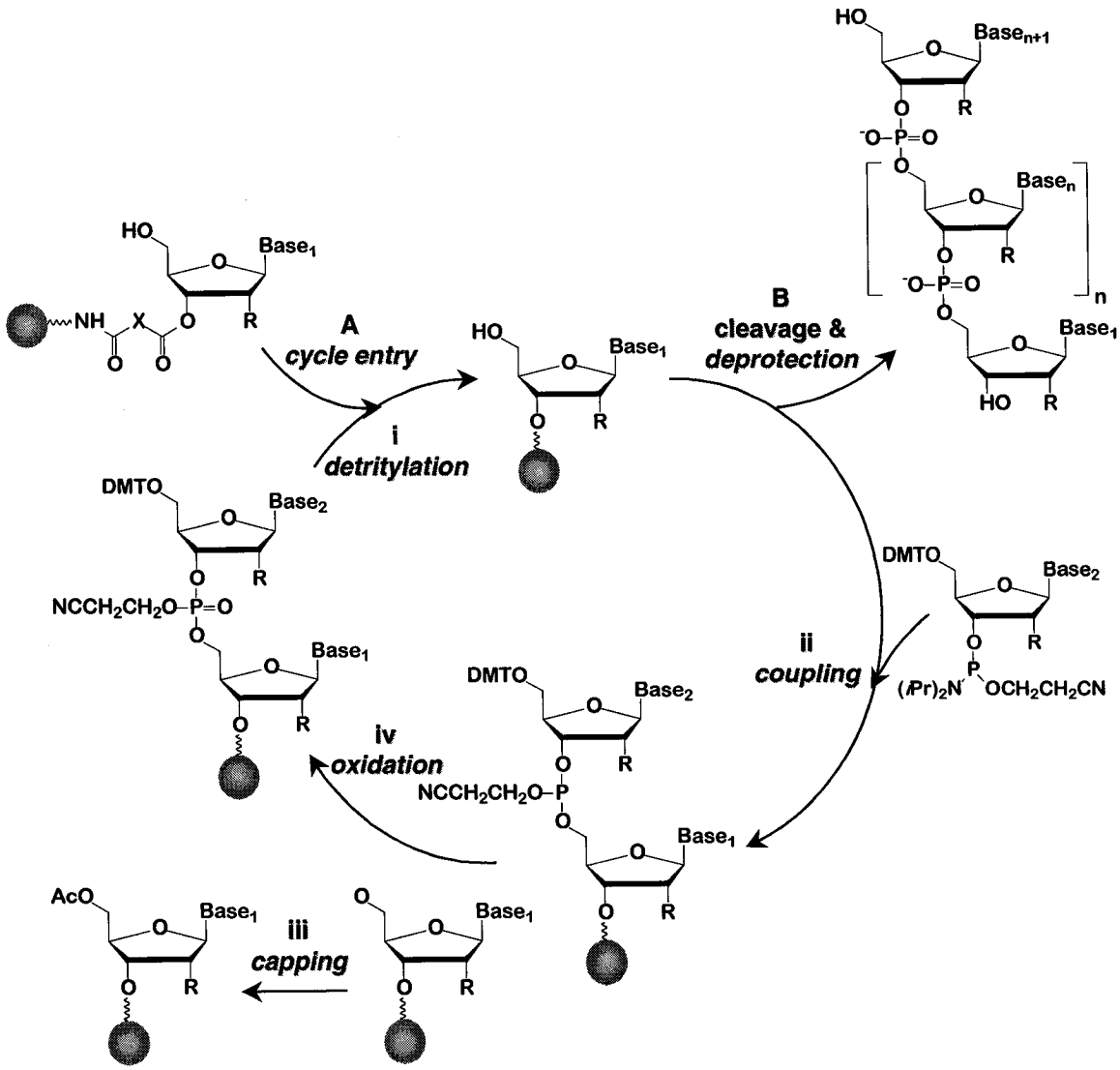


Figure 1.11: Illustration of automated solid-phase DNA and RNA oligonucleotide synthesis *via* phosphoramidite synthons and phosphite triester chemistry. (A) *Cycle entry* with a support-tethered and suitably protected nucleoside; (i) *detritylation* using 3% TCA in 1,2-dichloroethane; (ii) *coupling* of the free 5'-OH to a nucleoside 3'-phosphoramidite using an activating reagent (*e.g.* tetrazole); (iii) *capping* of unreacted 5'-OH groups using acetic anhydride; (iv) *oxidation* of the phosphite triester with iodine solution; (B) *Cycle exit* along with cleavage and deprotection of oligonucleotide chain after n+1 cycles.

(TBAF)[172,173] or triethylammonium trihydrofluoride (TREAT-HF)[182] is required for the complete deprotection of RNA oligonucleotides (2'-desilylation). Importantly, this "neutral" treatment is conducted as a final step since exposure to a very basic environment after 2'-OH release could potentially result in chain cleavage.[183]

1.7. CHEMICAL SYNTHESIS OF BRANCHED AND LARIAT NUCLEIC ACIDS

The discovery of branched RNA in 1983[99], later determined to be the intronic lariat intermediates formed during messenger RNA splicing[93,101], widened the eyes of many nucleic acid chemists. Almost immediately, the chemical synthesis of unique branched nucleic acids (bNAs) and oligonucleotides, reminiscent of the naturally occurring lariat species, commenced. The rapid availability of branched nucleic acids was thought to be of considerable value as such structures would be highly beneficial towards elucidating the mechanism of RNA splicing and lariat formation. Later, other bNA structures such as the "Y" or fork-shaped *trans*-splicing intermediates observed in parasitic nematodes[45,184], and prokaryotic multicopy single-stranded DNAs (msDNAs), containing an unusual branched structure of single-stranded DNA covalently linked to RNA[114,185], were also described. Regardless, the native branched molecules all contained the same inherent feature, namely a branchpoint nucleotide comprising vicinal 2',5'- and 3',5'-phosphodiester linkages. Early reports on the chemical synthesis of small bNAs containing tri- and tetranucleotides involved solution-based methods alone.[186-194] The particular synthetic challenge in producing such structures was in the devising tolerable orthogonal protecting groups, which would allow for the sequential introduction of vicinal 2'- and 3'-phosphodiester linkages at the branch point adenosine, specifically if diverse nucleotides were desired at the two positions. As such, it was clearly evident that introduction of a second phosphate residue, adjacent to an already incorporated internucleotidic phosphate (or phosphate derivative) at the branchpoint, required that this phosphate be weakly electrophilic, thereby imparting adequate resistance to isomerization reactions *via* participation of the vicinal hydroxyl group.[195,196] This excluded the use of internucleotidic *phosphotriesters*[197,198] since these intermediates are particularly vulnerable to nucleophilic attack by the neighboring hydroxyl under neutral, basic or acidic conditions, thereby resulting in phosphoryl migration, and potential chain cleavage.[189,195,197,198] Alternatively, *phosphodiester* internucleotidic linkages were shown to be reasonably resistant under a variety of conditions typically exploited for the selective removal of secondary hydroxyl protecting groups.[189,192]

The first reports on the solution-based synthesis of branched triribonucleotide diphosphates containing the native branched topology emerged in 1985 with concurrent reports from Ogilvie and co-workers[186] as well as Sekine and Hata.[188] Although intermediary phosphite triester linkages proved to be stable for introduction of simultaneous 2'- and 3'-phosphodiester bonds *via* phosphoramidite building blocks using Ogilvie's method, the *in situ* reaction of diverse 5'-phosphoramidites at the vicinal hydroxyl positions produced a mixture of regioisomeric branched species.[186] Kierzek and co-workers were the first to exploit the stable properties of phosphodiester linkages for the regiospecific creation of a branched tetraribonucleotide structure, $GA^{2',5'}(G)_{3',5'}C$, reminiscent of the native branched core in lariat introns.[189] Their novel methodology exploited phosphoramidite chemistry to assemble the $A_{3',5'}C$ dimer, which consisted of an internucleotide phosphotriester bearing the β-cyanoethyl protecting group as well as a 2'-TBDMS on adenosine. The β-cyanoethyl group was selectively eliminated under basic conditions, thus converting the highly reactive phosphotriester to the more stable phosphodiester. Cleavage of the 2'-TBDMS was then effectively accomplished in the absence of any undesired phosphoryl migration or chain cleavage, such that the branch could be extended from this position with a guanosine 5'-phosphoramidite, affording the branched trimer, $A^{2',5'}(G)_{3',5'}C$. Chain extension at the 5'-position ensued, resulting in the production of the tetrameric, branched nucleic acid. Other groups followed suit and reported alternative strategies, protecting groups and phosphorus chemistries for the introduction of vicinal phosphodiester linkages in small, branched nucleic acid structures.[191,192,199,200] A particularly interesting strategy for the synthesis of branched tri- and tetranucleotides emerged from the laboratory of Chattopadhyaya[194,201], and involved nucleotide assembly using H-phosphonate chemistry, initially introduced by Todd and Hall in 1957.[202] The technique relied on first converting the intermediate H-phosphonate in a diribonucleoside monophosphonate into a more stable phosphodiester adduct by oxidation, followed by selective cleavage of the neighboring hydroxyl protecting group. Previous efforts in assembling branched oligomers *via* this H-phosphonate strategy by Huss and co-workers failed miserably

owing to the fact that the internucleotide phosphonate linkage had not previously been converted to a more stable entity, prior to removal of the adjacent hydroxyl protecting group.[196]

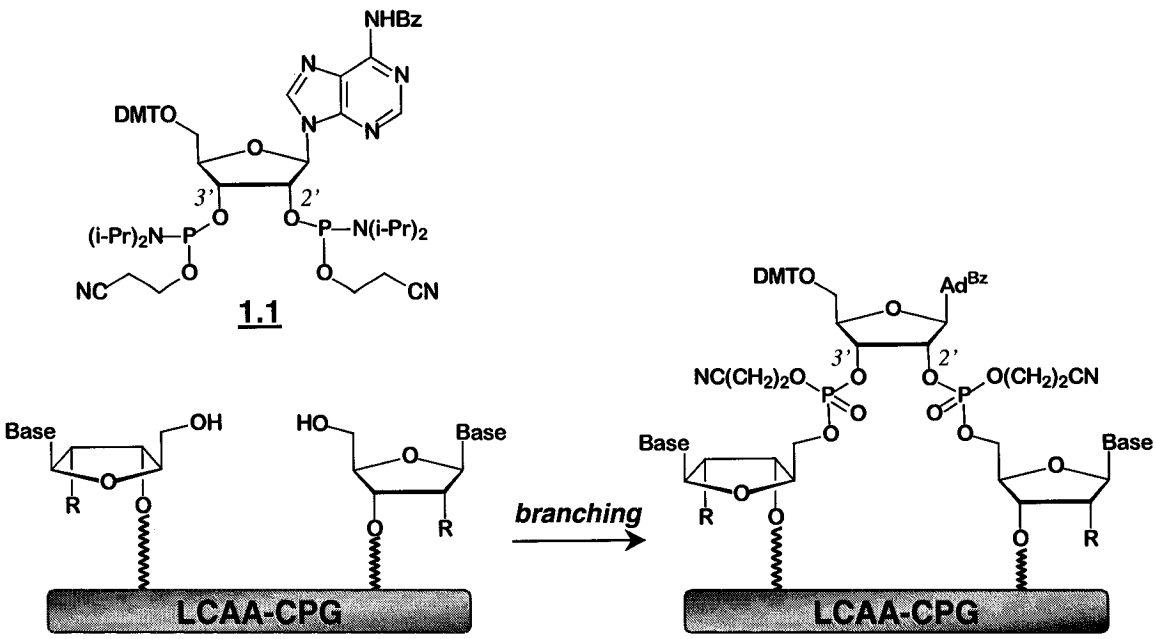


Figure 1.12: Convergent branched oligonucleotide synthesis developed by Damha *et al.* R = H in DNA; R = OTBDMS in RNA. i-Pr = isopropyl; Ad^{Bz} = N^6-benzoyladenine.

Although the approaches described for the creation of branched nucleic acids afforded moderate to high yields of small, branched structures, their multi-step synthetic methodologies limited their applicability toward the conception of larger and more biologically relevant branched oligonucleotides. It was the early solution-phase work by Ogilvie and co-workers that truly laid the foundation for the rapid, reliable and efficient solid-phase assembly of longer branched oligonucleotides by Damha and co-workers using phosphite triester chemistry.[186,187] Their novel strategy relied on the induction of a bifunctional phosphoramidite containing 2'- and 3'-functionalities, namely a bis-phosphoramidite (Compound **1.1**; **Figure 1.12**). Effective synthesis of this building block was accomplished using known, and well-established methods for the protection and activation of phosphoramidite moieties.[203] The key advantage of this methodology was that simultaneous introduction of the 2'- and 3'-phosphate groups at a branchpoint could

now be afforded without need for orthogonal deprotection steps and diverse phosphate chemistries at the two positions. In solution, treatment of the bis-phosphoramidite **1.1** with excess nucleoside bearing a free 5'-hydroxyl, in the presence of an activation reagent such as tetrazole[176,177] and subsequent oxidation, a variety of branched trinucleotide diphosphates were obtained.[187] As the 2'- and 3'-phosphodiester linkages were established concurrently, the method was not regiospecific, and only allowed for the presentation of identical nucleotides at the branchpoint. Regardless, the branched topology was introduced and maintained with few steps, and they envisioned that this strategy could be readily adapted to the synthesis of larger branched molecules. Indeed, in 1987 Damha and Zabarylo chronicled the first successful solid-phase synthesis of both DNA and RNA oligonucleotides containing homopolymeric stretches of dT_5 or rU_5 linked to an adenosine branchpoint (*i.e.* $A^{2',5'}(T_5)_{3',5'}T_5$ or $A^{2',5'}(U_5)_{3',5'}U_5$).[204,205] The *convergent* methodology exploited the tethering of two adjacent, CPG-bound, linear oligonucleotides containing free 5'-hydroxyl moieties with the bis-phosphoramidite **1.1**, to produce V-shaped nucleic acids containing identical sequences at both the vicinal 2' and 3'-positions off the branchpoint (**Figure 1.12** and **1.13**). Subsequent addition of commercially-available phosphoramidite building blocks at the apex of the V-structures afforded Y-shaped or "forked" oligonucleotides similar to the extended branch core in native lariats as well as the structures of *trans*-splicing intermediates (**Figure 1.13**).[45,184] The efficiency of the branching reaction proved to be highly dependent on both the extent of derivatization of the CPG with support-bound nucleosides (*i.e.* degree of loading) as well as on the concentration of bis-phosphoramidite utilized.[204,205] Optimal branching was obtained on supports containing higher loadings (*i.e.* 30-50 μmol/g of CPG), and with a dilute concentration of bis-phosphoramidite (*i.e.* 0.02-0.03 M). High CPG-loadings ensured that the reactive 5'-hydroxyls were in close enough proximity to one another for ample tethering with the bifunctionalized reagent (**Figure 1.12**). A low bis-phosphoramidite concentration prevented the formation of mostly "extended" linear isomeric failure sequences, which arise when the reagent reacts with only one of two support-bound linear strands. Compatibility of this methodology with silyl-phosphoramidite synthons, typically used in RNA oligonucleotide synthesis, allowed

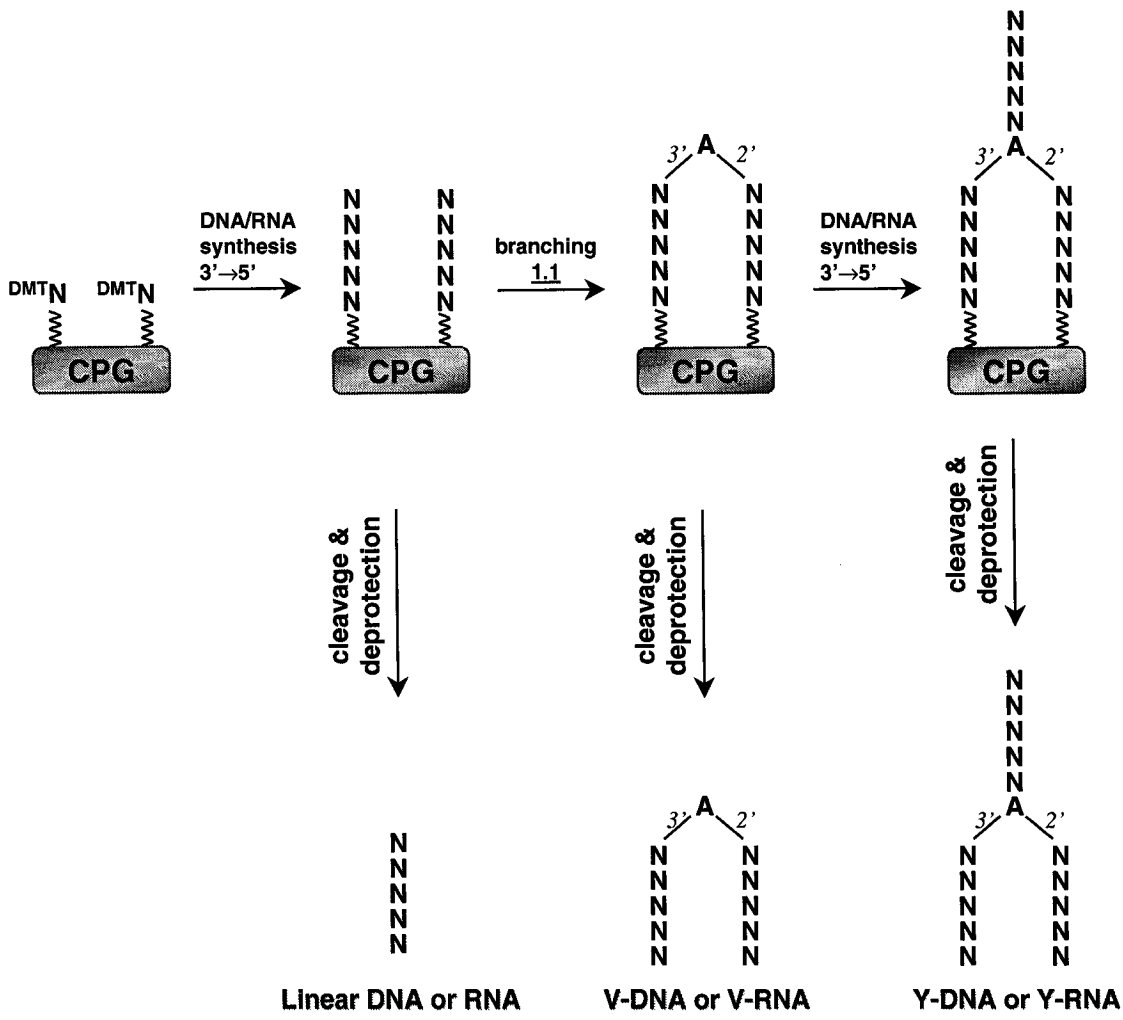


Figure 1.13: Convergent automated synthesis of V-shaped and Y-shaped branched DNA and RNA oligonucleotide structures on solid-support. The branchpoint is intoduced *via* the adenosine bis- phosphoramidite synthon (**1.1**). N=any nucleotide.

for the subsequent creation of branched RNA oligonucleotides containing mixed base composition, similar to the intronic consensus sequences inherent to RNA lariats.[206] Hyperbranched or "dendritic" DNA molecules up to 87-nucleotides have also been successfully assembled using this strategy by repeating the branching and chain elongation steps several times.[207,208] Such dendrimeric species have also been produced using a more efficient *divergent* strategy, in which the molecules are assembled in the 5'→3' direction rather than the normal 3'→5' direction typically used in automated synthesis.[209] Both the V-shaped and Y-shaped molecules have proven expedient for

studying triple-helix formation in branched oligonucleotides[210], as fiber-optic biosensors[211] as well as substrates for elucidating the sequence and stereochemical requirements at the branchpoint sugar for hydrolysis by the yeast and human debranching enzymes.[115,116,208,212,213]

By using a variation of the aforementioned approach, and an orthogonal deprotection step Damha and co-workers also reported the regiospecific synthesis of branched RNA molecules containing the wild-type consensus sequences found in RNA lariats by tethering two diverse linear RNA strands with the bis-phosphoramidite **1.1**.[214,215] Although branching resulted in the formation of four isomeric branched compounds in roughly 1:1:1:1 ratios, the oligomers were easily resolved by capillary electrophoresis (CE) and the approach provided the simplest method for preparing bRNAs with wild-types consensuses at the time. Similarly, Sproat and co-workers also reported novel methods toward the regiospecific solid-synthesis of small to medium-sized branched RNA molecules[216,217] in addition to branched DNA/RNA chimeras.[218] Although remarkably elegant, their strict methodologies necessitated numerous orthogonal protecting groups as well as the in-house synthesis of a complex set of phosphoramidite building blocks and branchpoint introduction synthons, making this approach highly labor intensive. The first fully-automated, regiospecific solid-phase synthesis of branched DNA molecules of arbitrary length and sequence composition using only commercially obtainable phosphoramidite synthons was accomplished by Jesper Wengel's group.[219,220] Their method relied on the stepwise introduction of various residues off the unprotected 5'- and 2'-positions of a branchpoint **X** using diverse concentration ranges and coupling times of phosphoramidites. The widespread use of this method would nonetheless be limited, as much user-specific optimization would be required. By looking back to the solution-based phosphodiester-intermediate chemistry of Kierzek and co-workers[189] for the introduction of regiospecific branchpoints, Braich and Damha were successful in adapting this method to the regioselective solid-phase synthesis of branched DNA[221] and branched DNA/RNA chimeras similar to the msDNA molecule of the prokaryote, *Myxococcus xanthus*.[222] Specifically, their mostly automated procedure

relied on commercially available phosphoramidite building blocks, but most significantly on the chemoselective removal of an alkylsilyl protecting group (*i.e.* 2'-TBDMS) under neutral conditions to afford chain extension from the 2'-hydroxyl of the branch with inverted 5'-phosphoramidites (see Chapter 3; **Figure 3.10**).

Although numerous successful strategies have been reported for the syntheses of small to medium-sized branched oligonucleotides *via* both solution and solid-phase methods, exceedingly few reports have materialized concerning the synthesis of lariat RNA molecules, specifically containing a circularized RNA with a 3'-tail extension, as seen in the native structures. The major challenge in creating such topologically unique molecules has been in devising methods to afford selective circularization of oligonucleotide sequences between the 5'-end of the branchpoint and the 2'-hydroxyl functionality. The first successful reports on the synthesis of a lariat RNA, bearing the natural branch core sequences and architecture, were put forward by Chattopadhyaya and co-workers.[223,224] Synthesis was conducted in solution and required the production of numerous non-standard nucleoside building blocks, which were laborious to prepare. Additionally, the solution-based method limited the size of the lariat species that could be assembled, and as such, the loop sizes obtained were only 3 to 5 nucleotides long. Interestingly however, their examination of the structure-dynamics relationship of such small lariat species through NMR techniques allowed them to observe self-cleavage properties in some of their molecules, comparable to the ribozyme activity seen in the RNA-hammerhead.[225] Alternatively and most recently, Reese and co-workers described the synthesis of branched cyclic oligoribonucleotides (*i.e.* small RNA lariats) using the solution-based H-phosphonate coupling and an extensive array of non-standard building blocks bearing orthogonal protecting groups.[226] Again, the largest lariat produced consisted of only 5 nucleotides, with a 3-nucleotide loop structure, which is not at all reminiscent of the native lariats.

1.8. THESIS OBJECTIVES

The natural occurrence of lariat RNA as biological intermediates during pre-mRNA splicing has prompted many groups to investigate facile synthetic methodologies for creating such compounds, or compounds reminiscent of the lariat itself (*i.e.* branched oligonucleotides). While many groups have been successful at constructing branched nucleic acids, reports on lariat-shaped molecules are scarce. The implications of a facile and efficient method for assembling such entities could lead to remarkable biochemical investigations into why evolution selected this novel topological form, as well as studying branchpoint selection during pre-mRNA splicing, or substrate requirements of the unique lariat debranching enzyme, a specific 2'-phosphodiesterase. With a vested interest in developing a simple and novel solid-phase based approach for synthesizing lariat oligonucleotides, we looked towards exploiting the branched topology of convergently synthesized Y-shaped DNA molecules, followed by cyclization of the 5'- and 2'-branch extensions in Chapter 2. Specifically, the differential rates of cleavage of two base-labile oligonucleotide tethers, the hydroquinone-O,O'-diacetate (*Q-linker*) and the succinate (*S-linker*) linkages, will be investigated for their propensity to selectively release the 3'-hydroxyl of a support-bound branched oligonucleotide for cyclization with an activated 5'-phosphoramidite present on the same molecule. Furthermore, the effects of increasing the CPG loading, thereby reducing the distance between adjacent oligonucleotides, on Y-DNA cyclization will also be explored.

In an effort to introduce a regioselective and high-yielding approach to lariat oligonucleotide synthesis of arbitrary base composition, the template-mediated syntheses of lariat-DNA and RNA oligonucleotides will be detailed in Chapter 3. The utility of branched DNA and RNA precursors (*i.e.* Y-shaped) containing regions of self-complementarity (*i.e.* dumbbells) or capability of associating with a complementary oligonucleotide splint, for the chemical ligation of medium-sized lariats, will be highlighted. Parameters for the successful cyclization of branch extensions and new phosphodiester bond construction during lariat synthesis will also be described.

Previous work from our laboratory has demonstrated that RNA-hairpin molecules containing the unique UUCG structural loop motif are potent and selective inhibitors of the RNase H activity of Human Immunodeficiency Virus Type 1 Reverse Transcriptase (HIV-1 RT).[227] The high thermal stability, similarity in structure to hairpin RNA, and nuclease resistance of dumbbell nucleic acids prompted us to evaluate the inhibitory potential of RNA dumbbells towards the RNase H functionality of HIV-1 RT in Chapter 4. Methods for the successful and efficient ligation of the RNA dumbbells using purely chemical means (*i.e.* cyanogen bromide) will also be detailed.

In Chapter 5, the utility of branched oligonucleotides containing either DNA or RNA nucleotides, alternate sugar conformations at the branchpoint (*i.e.* ribose *vs* arabinose), as well as modified residues in the branch extensions (*i.e.* RNA *vs* 2'-O-methyl RNA) for studying branchpoint recognition events during pre-mRNA splicing will be revealed. Additionally, the use of unnatural nucleotide residues at the 3'-termini of the branched structures will be explored as a means of stabilizing the oligonucleotide structures against omnipresent exonuclease activity present in both the yeast and mammalian nuclear extracts. Furthermore, Chapter 5 will highlight preliminary results on the *in vitro* modulation of alternatively spliced isoforms of Bcl-x, a gene directly related to controlling cellular apoptosis, using linear and branched antisense oligonucleotide constructs containing spliceosomal binding overhangs. Specifically, the propensity of shifting 5'-splice site selection from the anti-apoptotic Bcl-x_L to the pro-apoptotic Bcl-x_S using such antisense constructs will be investigated.

CHAPTER 2: SOLID-PHASE SYNTHESIS OF LARIAT DNA ON A MIXED LINKER (Q/S-LINKER) SOLID-SUPPORT

2.1. INTRODUCTION

"Lariat" RNAs have been of specific interest to both the chemical and biological communities since their discovery two decades ago as splicing intermediates (branched RNA introns) in the post-transcriptional biosynthesis of mRNA.[94,99,101] Since that time, a number of research efforts have been devoted to synthesizing oligonucleotides related to these structures for studying the mechanism and catalytic factors required for lariat formation and metabolism during RNA splicing. The inherent branched architecture of these molecules is ascribed to a highly conserved branch point adenosine moiety bearing vicinal 2',5' and 3',5' internucleotidic phosphodiester linkages. Synthetic branched analogues (Y-shaped molecules) of these unique biomolecules have been particularly useful for probing the structural requirements of branch point recognition during splicing[228] and of lariat debranching enzymes.[115] In addition, they have been used as haptens for isolating antibodies specific to branched nucleic acids (bNAs)[229], as potential agents for controlling gene expression through triple-helical complex formation[210,219,220,230,231], and as nucleic acid biosensors.[211] Many groups, including our own, have been successful in chemically synthesizing bNAs (*i.e.* Y-shaped oligonucleotides) *via* both solution[186-190,192,200] and solid-phase[204,206,214,216,218-221] strategies, however few successful endeavors have described synthetic tactics towards authentic lariat oligonucleotide structures (*i.e.* circular RNA with a tail), demonstrating that their synthesis has been a major challenge. RNA "mini"-lariat syntheses have been reported;[223,224,226] nevertheless, the methodologies require the use of non-standard nucleoside building blocks, are extensively time consuming (many chromatographic purification steps), and are restricted to the synthesis of small lariats (di-, tri-, tetra- and pentameric loops) which are non-representative of the naturally occurring biomolecules. Furthermore, since the cyclization reactions were conducted in solution, high-dilutions of reactant were requisite in order to prevent

inevitable dimerization reactions and other high molecular weight polymers. Efforts toward the solid-phase synthesis of lariat oligonucleotides were further investigated in our research laboratory.[215] A solid-phase synthetic approach to lariat oligonucleotide synthesis would be highly advantageous over a solution-based method, since the length of the oligonucleotide that could be synthesized would not be as restricted. The developed approach involved the incorporation of a branchpoint dinucleoside triphosphate [$A^{2',5'}$(p*Up*)$_{3',5'}$p: p=phosphodiester; p*=phosphotriester] to the 5'-hydroxyl terminus of a support-bound DNA using the phosphotriester coupling chemistry developed by Letsinger and Ogilvie.[150,151] Chain extension from the 5'-terminus of the adenosine branch point ensued, followed by selective removal of the uridine-3'-phosphate cyanoethyl protecting group, and cyclization between the 5'-terminal hydroxyl of the grown strand and the resulting 3'-phosphodiester moiety of the 2'-uridine residue with the coupling reagent MSNT (1-(mesitylenesulfonyl)-3-nitro-1,2,4-triazole). Although this synthetic strategy has been successfully utilized in the solid-phase condensation of cyclic oligonucleotides[232], dimerization between two adjacent oligonucleotide strands was a key problem, owing to the inaccessibility of the reactive phosphotriester moiety of the uridine nucleotide by the intramolecular 5'-hydroxyl unit.

2.2. PROJECT OBJECTIVE

Unlike the synthesis of circular or branched oligonucleotides, lariat synthesis is highly complicated by the fact that additional steps and protecting groups are typically required for chemoselective introduction of a tail extension in the molecule. In an effort to overcome the inaccessibility of cyclization with branched nucleic acid precursors[215], and to devise of a solid-phase strategy for lariat synthesis using mainly commercially available phosphoramidite building blocks, we looked towards our well established protocol for the convergent synthesis of branched oligonucleotides[204,206,214,233], and the differential rates of cleavage of two tethering oligonucleotide-CPG linkers. Mixed linker CPG resins (1:1 ratio of linkers) were prepared using an equimolar ratio of both a standard 3'-succinylated thymidine nucleoside (**2.2**) and a novel 3'-hydroquinone-O,O'-

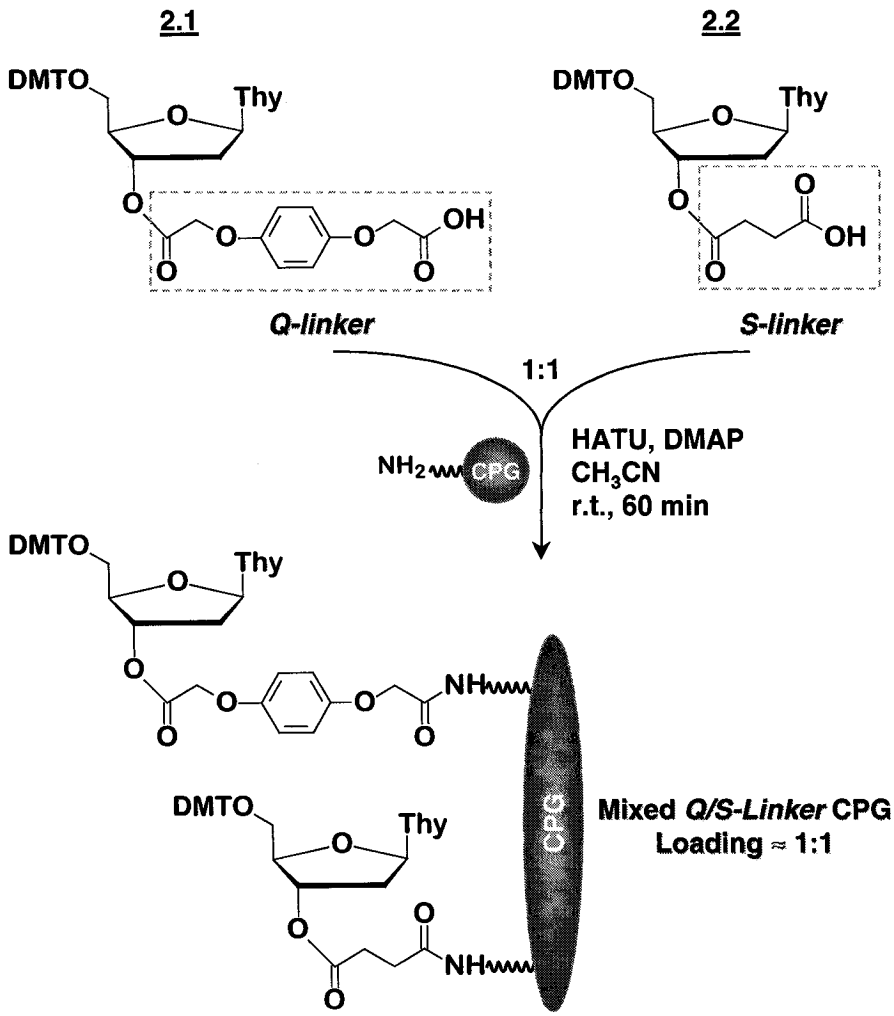


Figure 2.1: Synthesis of mixed Q/S-linker CPG in an approximate 1:1 ratio. Q-linker = hydroquinone-O-O'-diacetyl linker; S-linker= succinyl linker.

diacetyl thymidine nucleoside (**2.1**) developed by Pon and co-workers (**Figure 2.1**).[167,168] Branched Y-DNA **2.4** was convergently synthesized as demonstrated in **Figure 2.2** on the mixed Q/S-linked CPG using the suitably protected adenosine bis-phosphoramidite monomer (Compound **1.1**)[187], notably the only phosphoramidite building block that needed to be prepared in-house. Branching reactions were conducted on diverse loadings of nucleoside support (20-83 μmol/g) in order to clarify how changes in distance between oligonucleotide strands effected both the branching and cyclization reactions. Subsequent to Y-DNA synthesis, the 5'-terminal hydroxyl was phosphitylated

with β-cyanoethyl-(N,N-diisopropylamino) phosphochloridite reagent, followed by selective hydrolysis of the labile Q-linker tether from the solid support, thereby releasing a free 3'-hydroxyl unit (**Figure 2.3**). Using standard phosphoramidite coupling, the reactive 5'-phosphoramidite and 3'-hydroxyl could then be affixed, affording an intramolecularly cyclized DNA lariat structure that is still attached to the CPG surface through the more robust succinyl linkage. As a final step, treatment of the cyclized lariat structure under standard deprotection conditions would completely unmask any phosphate and base protecting groups with concomitant cleavage of the oligonucleotide-succinyl linkage from the CPG surface (**Figure 2.3**).

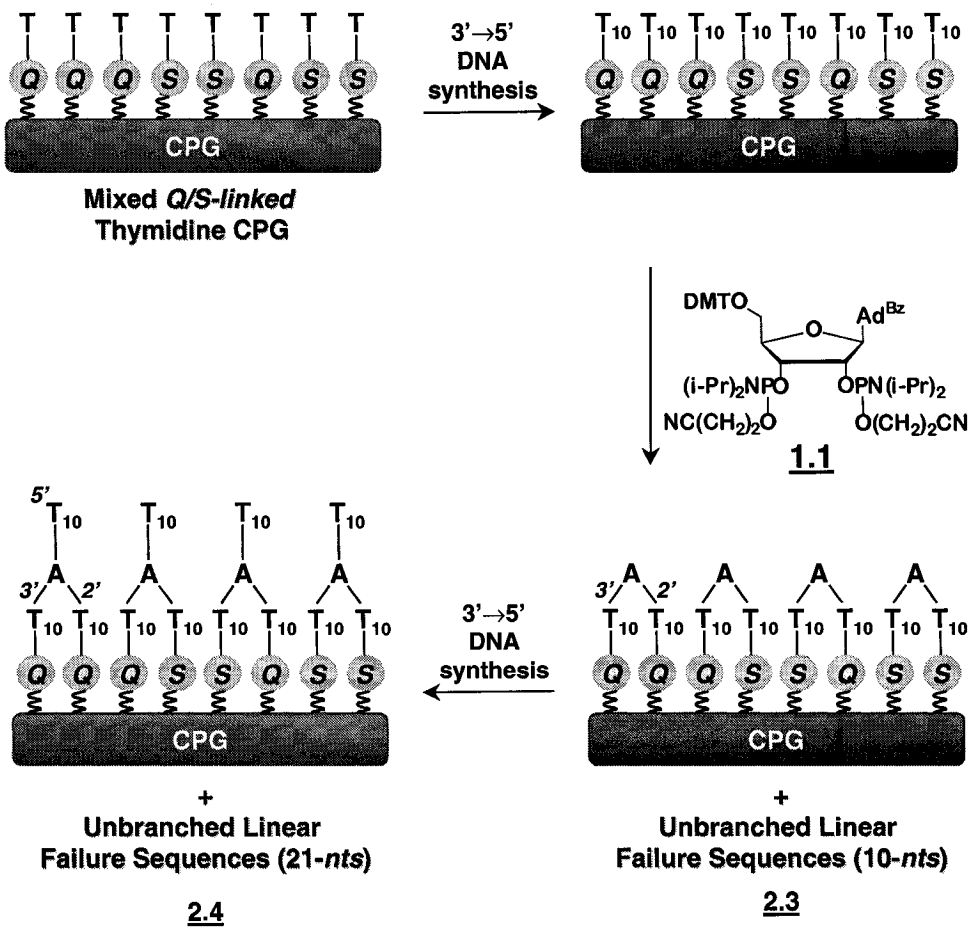


Figure 2.2: Convergent synthesis of branched DNA, $T_{10}A^{2',5'}(T_{10})_{3',5'}T_{10}$ on mixed Q/S-linked CPG. The synthetic methodology results in the formation of two predominant failure products: (1) unreacted 10-nt failure sequence, T_{10}, resulting from the failure of the bis-phosphoramidite **1.1** to react with the support-bound T_{10} strands and (2) isomeric 21-nt failure sequences $[T_{10}A^{2',5'}(p)_{3',5'}T_{10} + T_{10}A^{2',5'}(T_{10})_{3'}p]$ resulting from the reaction of the bis-phosphoramidite with only one T_{10} strand, followed by chain extension.

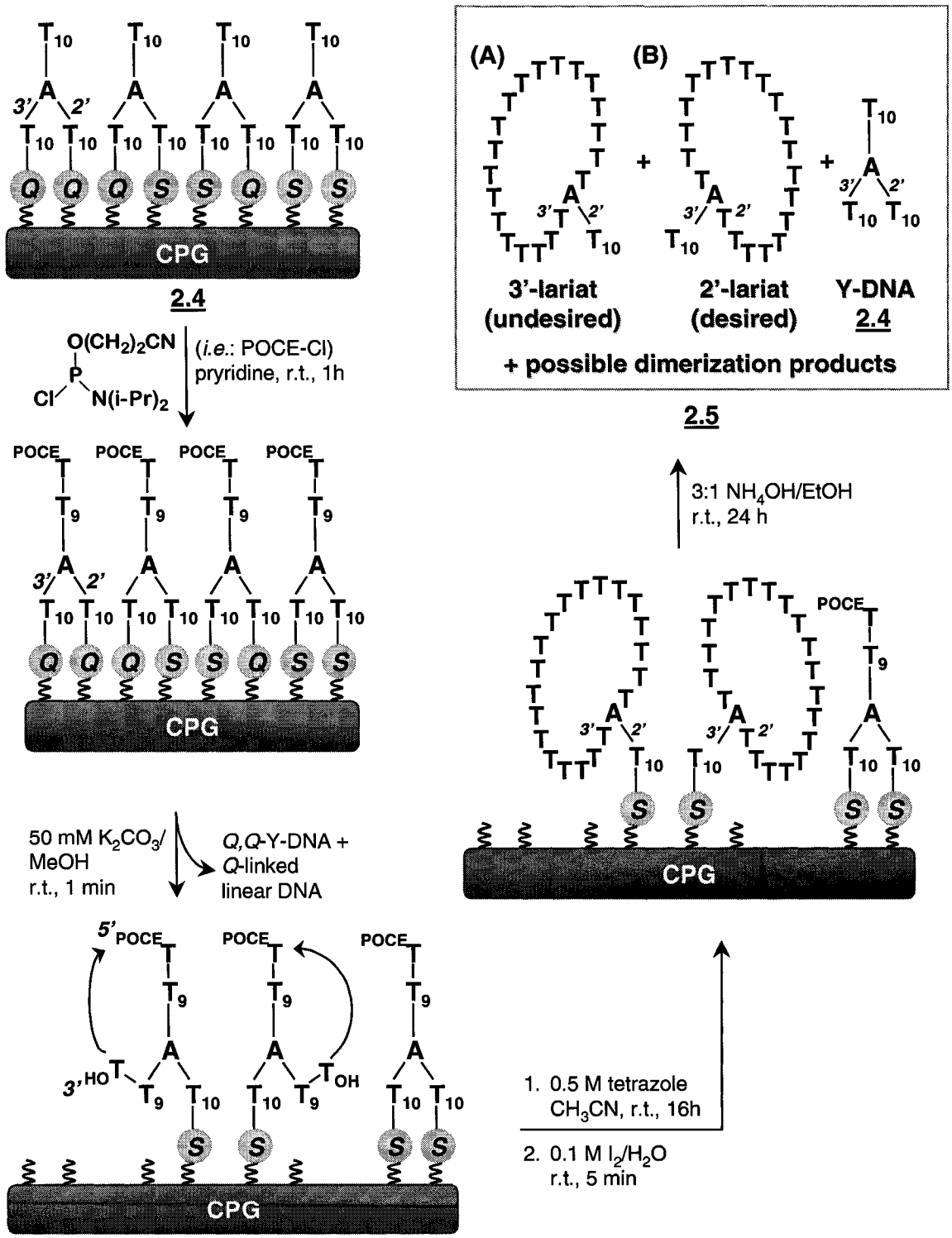


Figure 2.3: Schematic representation of the cyclization of branched DNA $T_{10}A^{2',5'}(T_{10})_{3',5'}T_{10}$ on mixed Q/S-linked CPG. POCE=cyanoethyl phosphite protecting group.

2.3. PHOSPHITYLATION OF THE 5'-HYDROXYL OF A SUPPORT BOUND NUCLEOSIDE: SYNTHESIS OF A MODEL THYMIDINE DIMER

Key to the success of lariat cyclization was the selective introduction of a reactive phosphoramidite moiety at the 5'-terminus of the support-bound branched oligonucleotide strand (**Figure 2.3**). At the time that these experiments were being conducted, no known examples of the phosphitylation of a nucleoside or oligonucleotide bound to the CPG surface were evident to us. We rationalized that the free 5'-hydroxyl group could be selectively phosphitylated using the highly reactive, β-cyanoethyl-(N,N-diisopropylamino) phosphorochloridite, a reagent commonly used in the preparation of nucleoside phosphoramidites for solid-phase oligonucleotide synthesis. In an attempt to demonstrate that this reaction could be fruitfully used in the cyclization of oligonucleotides still tethered to the CPG surface, a model system was devised whereby a support-anchored thymidine nucleoside was phosphitylated using the phosphorochloridite reagent, and then coupled to the free 3'-hydroxyl of a 5'-protected thymidine nucleoside, producing a thymidine dimer (**Figure 2.4**). Prior to phosphitylation, the 5'-dimethoxytrityl (DMT) protecting group was removed by treatment of the CPG with 3% (w/v) trichloroacetic acid in dichloroethane (3% TCA/DCE) on the DNA synthesizer. The support was suspended in CH_3CN and treated with 5, 10 or 100 equivalents (based on a 1 μmol synthesis) of chlorophosphoramidite reagent. Pyridine was added as a general base to quench any hydrochloric acid formed during the reaction. After frequent washing and thorough drying of the support to remove any residual moisture, which would compete with the incoming nucleophile, it was resuspended in CH_3CN and reacted with an excess of 5'-dimethoxytrityl thymidine nucleoside with 1H-tetrazole used as an activating reagent. Subsequently, the phosphite triester intermediate was oxidized *in situ* to the more stable phosphotriester with an iodine/water solution, washed thoroughly and dried. The extent of coupling was determined by trityl cation (DMT^+) release from the 5'-terminus of the resultant dimer. Elimination of the cyanoethyl phosphate protecting group and cleavage of the dimer from the CPG surface was effected under standard ammonia deprotection conditions.

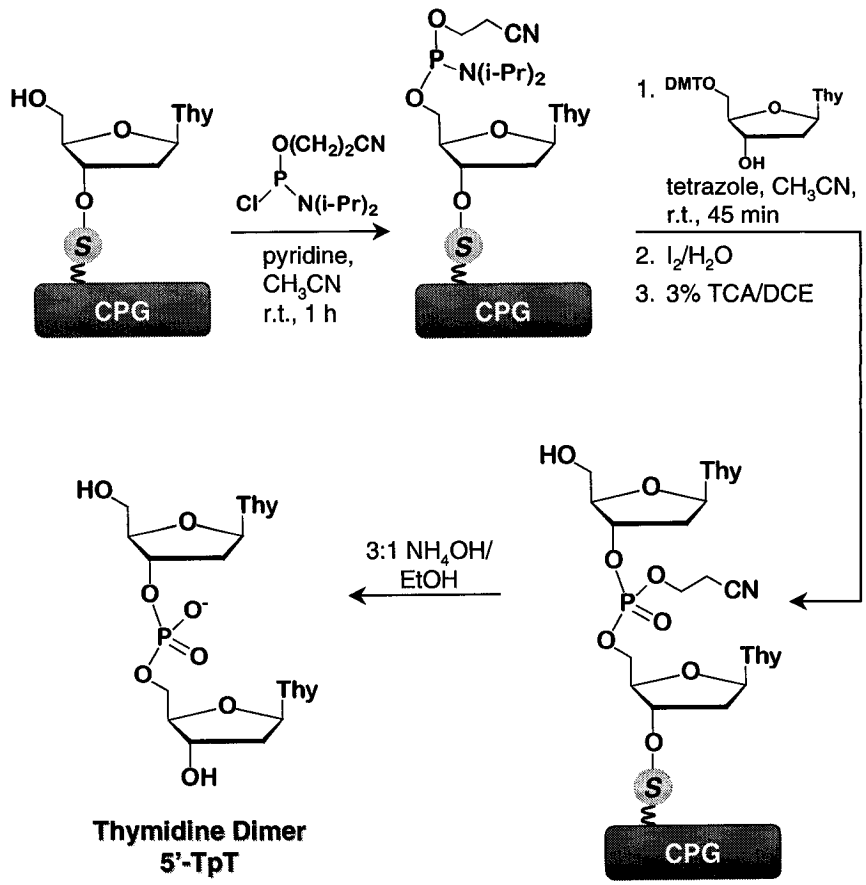


Figure 2.4: Synthesis of a model thymidine dimer on solid-support *via* phosphitylation and coupling to a support-bound nucleoside.

Thymidine dimer product formation was ascertained by anion-exchange HPLC (**Figure 2.5**) and denaturing PAGE analysis on a highly crosslinked gel (not shown). Analysis of the HPLC profiles revealed that dimerization on the solid-support was indeed successful, with a predominant amount of dimer formed when 10 equivalents of chlorophosphoramidite reagent were used to phosphitylate the 5'-terminus. Two side products corresponding to the unreacted support-bound thymidine nucleoside (dT; t_R=2.7 min) and the hydrolyzed 5'-phosphoramidite (*i.e.* 5'-thymidine monophosphate; 5'-p(dT); t_R=19 min) were also evident, yet in much smaller proportions. When a significant excess of the phosphitylating reagent was used (*i.e.* 100 eq) the dimerization reaction occurred with only *ca.* 50% efficiency. This was surprising, as an appreciable excess of the chlorophosphoramidite would be expected to completely phosphitylate the 5' -termini, resulting in a substantial enhancement in the amount of dimer produced. Nonetheless, the

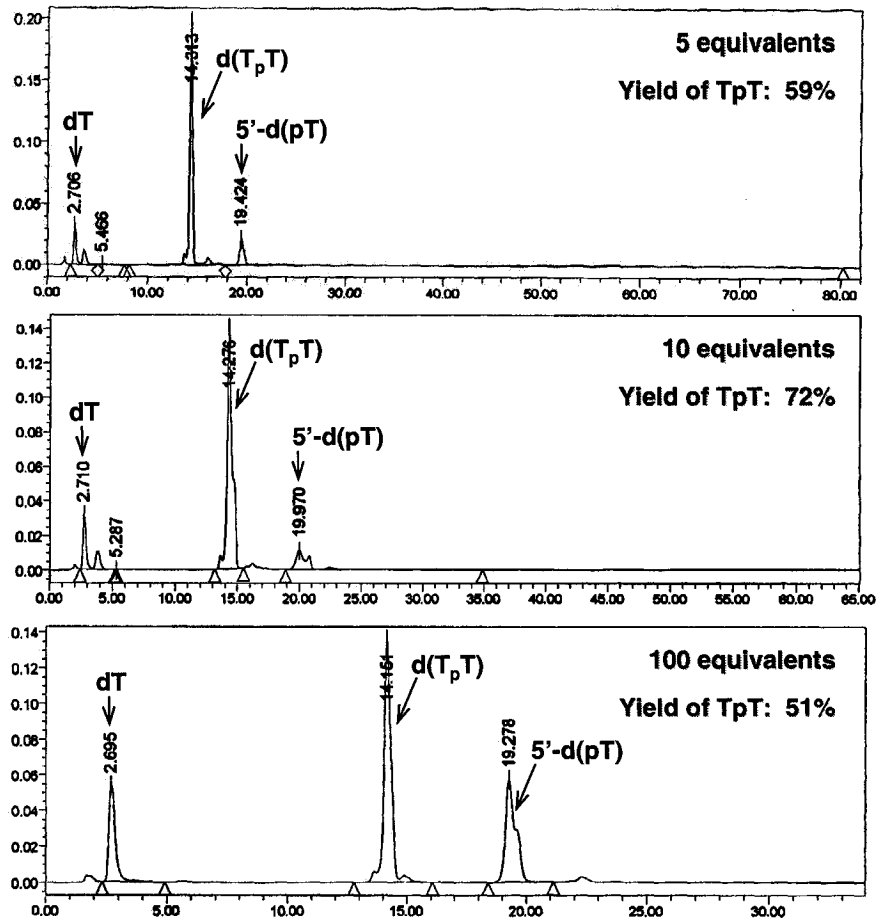


Figure 2.5: Anion-exchange HPLC chromatogram of the reaction products obtained from the synthesis of a model thymidine dimer on solid-support using 5, 10 and 100 equivalents of chlorophosphoramidite reagent. HPLC conditions: Protein-Pak DEAE-5PW (75 × 7.5 mm) column; Buffer A: ddH_2O; Buffer B: 0.2 M $LiClO_4$; Gradient: 0-50% Buffer B over 60 min. T= 50°C.

ensuing reactions were conducted only once, and it is likely that this was just an artifact of the reaction conditions at the time (*i.e.* increased moisture). It was later revealed to us, that 5'-phosphitylation of a support-bound oligonucleotide using the chlorophosphoramidite reagent had been investigated prior to our studies.[234] This invention specifically described methods for incorporating radiolabeled nucleosides at one or more predetermined sites within an oligonucleotide, through the activation of a support-bound, nascent strand with an activating reagent, such as the chlorophosphoramidite described herein. The activated oligonucleotide is then contacted with a 5'-protected, radiolabeled mononucleoside bearing a free 3'-hydroxyl, resulting in

the condensation of the nucleoside and phosphitylated oligonucleotide strand. The conditions utilized for phosphitylation, activation and coupling of the 5'-phosphoramidite/3'-hydroxyl termini in the aforementioned study were nearly identical to those utilized in the dimerization experiments. Particularly, a 30-fold excess of chlorophosphoramidite reagent was sufficient to adequately phosphitylate the 5'-termini of the CPG-tethered strand. The dimerization experiments clearly established that phosphitylation on support and successive activation and coupling to a free hydroxyl moiety is a plausible route to phosphodiester linkage formations and cyclization of the desired lariat structure.

2.4. SYNTHESIS OF LARIAT DNA ON MIXED Q/S-LINKED CPG

2.4.1. Synthesis of Branched DNA on Q/S-Linked CPG

Selection of the appropriate mixed CPG-anchoring system in this study necessitated that the two linker arms be resistant to the conditions utilized during solid phase synthesis, 5'-hydroxyl activation, and that they each demand distinctive conditions for cleavage of the oligonucleotide-linker bond. Knowing this, the succinyl (*S-linker*) and hydroquinone-O,O'-diacetyl (*Q-linker*) linker arms were selected (**Figure 2.1**). Nucleoside-3'-succinate linkages are the norm in solid-phase synthesis for appending the terminus of a nascent oligonucleotide to the solid support. Cleavage of this robust linkage is typically conducted in tandem with removal of any phosphate and base protecting groups on the oligonucleotide strand using aqueous ammonium hydroxide. Complete cleavage from the support typically requires at least a 2-hour treatment with ammonia solution. Alternatively, the more labile *Q-linker* is fully cleaved from the support in less than five minutes under identical conditions.[167] The succinyl linkage, on the other hand, has demonstrated complete resistance to ammonolysis after 5 minutes time. Although Q-linker cleavage is rapidly realized with ammonium hydroxide, such conditions would also eliminate the β-cyanoethyl phosphate protecting group, possibly resulting in undesirable side reactions. Otherwise, a milder cleavage reagent, 50 mM K_2CO_3 in methanol, would

also effect 100% cleavage in approximately one minute without consequence on the succinyl tether, phosphate and base protecting groups.[167]

Mixed Q/S-linked CPG comprising various nucleoside loadings was derivatized using the suitably functionalized 5'-O-DMT-thymidine-3'-O-succinate (**2.2**) and 5'-O-DMT-thymidine-3'-O-hydroquinone-O,O'-diacetate (**2.1**) nucleosides (**Figure 2.1**). The appropriate nucleosides were prepared according to established methods[167,168,235] and their identities confirmed by TLC comparison to the known R_f values and FAB-mass spectrometry (Experimental section 7.2.2). Alternatively, the 3'-O-succinate was also available commercially (ChemGenes Corp; Watham, MA.). The functionalized thymidine monomers were mixed in equimolar amounts with the CPG, and in order to ensure that an approximate 1:1 ratio of the two tethers were bound to the support surface, a rapid derivatization method by means of the uronium coupling reagent HATU, was utilized.[236] The CPG loading of the individual linkers was established by independent measurements of the [Q+S]-linker loading and the S-linker loading alone. The [Q+S] loading was determined by standard treatment of a small amount of CPG with 3% (w/v) TCA in DCE and measurement of the trityl cation absorbance at 505 nm. Alternatively, the S-linker loading was established by first treating the [Q+S]-linked CPG with aqueous NH_4OH for 5 min to cleave the Q-linked nucleosides, treating the resin with 3% TCA solution and finally measuring the absorbance of trityl cation released. The Q-linked loading was easily determined by subtracting the S-linker loading from the [Q+S] loading.

Branched DNA **2.4** was synthesized on mixed Q/S-linked CPG of various loadings (20-83 μmol/g) using the convergent synthesis methodology previously developed in our research group (**Figure 2.2**).[204,233] The branched oligomers were assembled using the well-established phosphoramidite chemistry and reagents on an ABI 381 DNA synthesizer. Branching of two vicinal decathymidylic oligonucleotide chains was appropriately effected using a 5'-O-DMT-(N^6-benzoyl)-adenosine-2',3'-O-bis-(β-cyanoethyl)-N,N-diisopropyl phosphoramidite (Chapter 1; Compound **1.1**) to produce the V-shaped nucleic acid **2.3** (**Figure 2.2**). An adenosine unit was included at the branch

point given that branch formation in native lariats occurs almost exclusively at this residue.[102,237] Synthesis of this monomer was conducted using published methods which relied on the transient benzoylation of riboadenosine[171], dimethoxytritylation of the 5'-hydroxyl group[170,238] and subsequent bis-phosphitylation of the 2' and 3'-hydroxyl units with an excess of β-cyanoethyl-(N,N-diisopropylamino) phosphorochloridite.[187,233] The monomer was utilized at a dilute concentration of 0.03 M, thus ensuring that branch product formation prevailed, as higher concentrations of the bis-phosphoramidite favor the creation of unbranched (linear) isomeric failure sequences (**Figure 2.2**). Chain extension from the tethered, branched DNA resulted in the construction of a Y-shaped oligomer consisting of a 2'- and 3'-T_{10} strands linked to the branchsite and a 5'-T_{10} extension off the branchpoint adenosine. Owing to the condensation of two adjacent linear oligonucleotides with the adenosine bis-phosphoramidite synthons, the resultant bNAs contain identical "branches" connected *via* vicinal 2',5'- and 3',5'-phosphodiester linkages. This symmetry was nonetheless inconsequential since the two vicinal strands contained homopolymeric thymidine sequences. Oligonucleotide synthesis was conducted in the "trityl-off" mode, thereby certifying that the 5'-terminal DMT group had been effectively removed to reveal a reactive 5'-hydroxyl. The glass beads were meticulously dried with argon and left in a vacuum dessicator overnight to exclude traces of moisture.

A small amount of CPG was removed at this point in order to verify bNA formation, and the oligonucleotide fully deprotected using aqueous ammonium hydroxide, consequently cleaving both the Q and S-linkers, and eliminating the benzoyl and cyanoethyl protecting groups. Denaturing PAGE analysis of the branched product **2.4** clearly established an important relationship between CPG-loading and branching efficiency (**Figure 2.6**). Three predominant species are cleanly resolved on the gel corresponding to the full length bNA product **2.4** [$T_{10}A^{2',5'}(T_{10})_{3',5'}T_{10}$], the unbranched linear isomeric failure sequences [$T_{10}A^{2',5'}(p^*)_{3',5'}T_{10}$ and $T_{10}A^{2',5'}(T_{10})_{3',5'}p^*$] and the truncated T_{10} oligomer resulting from unsuccessful reaction of the dilute bis-phosphoramidite with the terminal 5'-hydroxyl of the nascent T_{10} strand. Separation of bNA **2.4** from failure sequences by PAGE is greatly facilitated by the relatively large molecular weight of this species and amplified

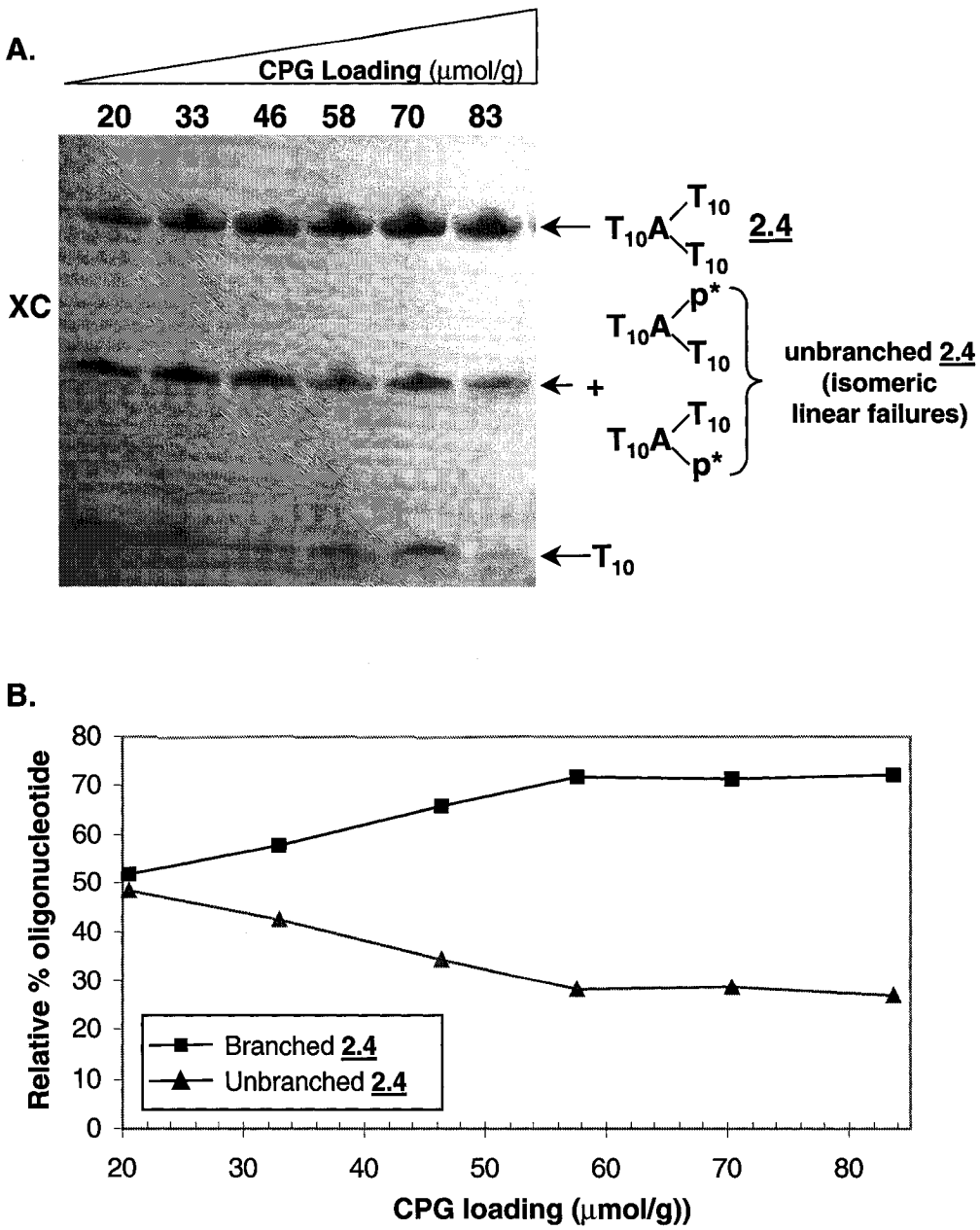


Figure 2.6: Effect of CPG loading (Q + S linkers) on the yield of Y-shaped bDNA, $T_{10}A^{2',5'}(T_{10})_{3',5'}T_{10}$ synthesized *via* the convergent branched DNA synthesis methodology. Panel A: PAGE analysis (20% acrylamide, 7M urea) of the amount of Y-shaped product (**2.4**) and isomeric linear failure sequences (Unbranched **2.4**) formed as a function of CPG loading; XC=xylene cyanol marker dye, p*=phosphate group. Panel B: Chart demonstrating the increase in the amount of the Y-shaped oligonucleotide (**2.4**) with increasing nucleoside-CPG loading. The chart also demonstrates the inverse relationship between the amount of extended isomeric linear failures (unbranched **2.4**) and the nucleoside loading on the CPG. The relative % oligonucleotide was determined *via* HPLC integration of the branched and unbranched oligonucleotide peak areas.

mass/charge ratio. Indeed, the branched nature of these structures imposes frictional effects as the molecules migrate through a cross-linked gel environment, thereby exhibiting retarded mobility compared to their linear correlatives of identical base-composition. Maximal branching and nominal isomeric failure sequence formation was noticeably achieved with higher CPG loadings (60-83 μmol/g; **Figure 2.6B**), corroborating the fact that decreasing the distance between adjacent oligonucleotide strands renders the 5'-hydroxyls in a more accessible position for the bridging reaction to occur.[204,205] Additionally, the crude branched oligomers were analyzed by anion-exchange HPLC (**Figure 2.7**), testifying that the full-length bNA was conveniently separated from its isomeric failure sequences owing to an increase in the number of charged phosphates in the structure. In fact, under the HPLC conditions utilized, the regioisomeric failure sequences were isolated into two distinct peaks of nearly equal intensity (**Figure 2.7**).

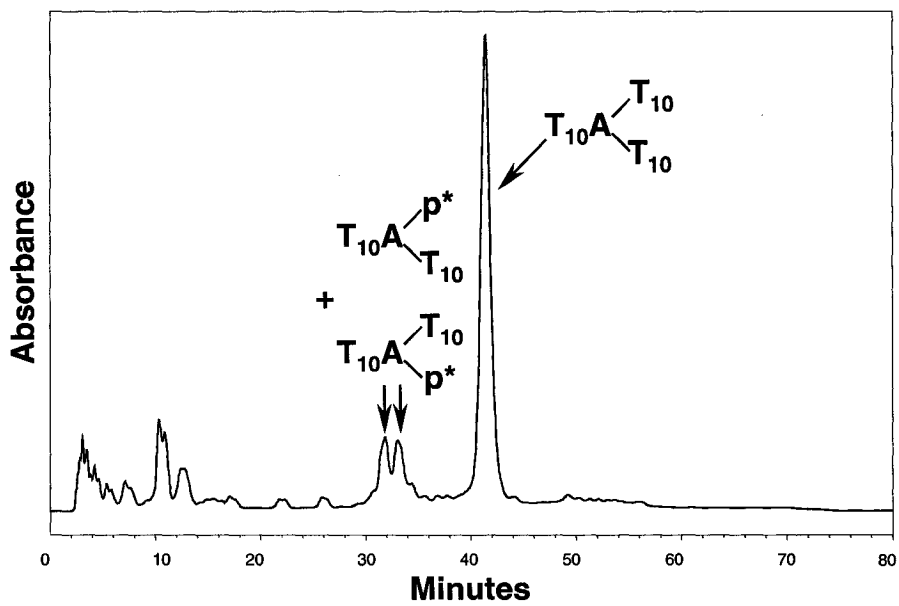


Figure 2.7: Anion-exchange HPLC chromatogram of crude **2.4** [$T_{10}A^{2',5'}(T_{10})_{3',5'}T_{10}$] synthesized on high-loading Q/S-linked CPG (83 μmol/g). HPLC conditions: Buffer A: ddH_2O; Buffer B: 1 M $LiClO_4$; Gradient: 10-20% Buffer B over 60 min; Column temperature=50°C.

2.4.2. Phosphitylation, Selective Q-Linker Cleavage and Cyclization of Branched DNA

The remaining, dry CPG containing the support-bound bNA **2.4** was selectively phosphitylated at its free 5'-OH terminus using an excess (50 equivalents) of the chlorophosphoramidite reagent (**Figure 2.3**). Subsequently, the oligonucleotide/Q-linker ester linkages were selectively cleaved by treating the glass beads with 50 mM K_2CO_3 in methanol for one minute. Previous ^{31}P-NMR evidence established that methanolic K_2CO_3 could accomplish rapid deprotection (5 min) of cyanoethyl-blocked phosphotriesters, indicating a possibility that the cyanoethyl group on the 5'-terminal phosphoramidite might be undesirably eliminated.[239] In order to exclude this possibility, a small amount of 5'-DMT-thymidine-3'-O-(β-cyanoethyl)-N,N-diisopropyl phosphoramidite (thymidine 3'-phosphoramidite) was treated with 50 mM methanolic K_2CO_3 for one minute, and the ^{31}P-NMR spectrum of the compound acquired (200 MHz; solvent: acetone-D_6). Elimination of the phosphite cyanoethyl group would expectedly cause the ^{31}P-NMR peak (peak position: 149 ppm) to move upfield in the spectrum. In effect, the spectrum revealed that the phosphoramidite remained unscathed, presumably the result of the particularly short deprotection time utilized (not shown). We further anticipated that this short reaction time (1 min) would not cause significant decyanoethylation of the internucleotide phosphotriester moieties. The methanolic supernatant was isolated and the intactness of the cleaved oligonucleotides assessed by denaturing PAGE (**Figure 2.8**). As expected, cleavage by the methoxide anion resulted in the release of any Q-linked linear, and Q,Q-linked bNAs (**Figure 2.3**), exhibiting an identical migratory pattern to those evidenced in the bNA synthesis gel (**Figure 2.6**). The CPG beads, now contained only the succinyl-tethered oligonucleotides. Furthermore, upon treatment with the methanolic K_2CO_3 solution, the Q-anchors of the 2'-Q, 3'-S-linked and 2'-S, 3'-Q-linked oligonucleotides were apparently cleaved to produce a free 3'-hydroxyl unit, while still tethering the oligonucleotide to the support *via* the S-linkage.

Prior to activation and cyclization of the support-bond oligonucleotides, it was imperative the CPG beads were utterly dry, so as to avert competing reactions between the 5'-phosphoramidites and any water molecules. The coupling reaction was further conducted

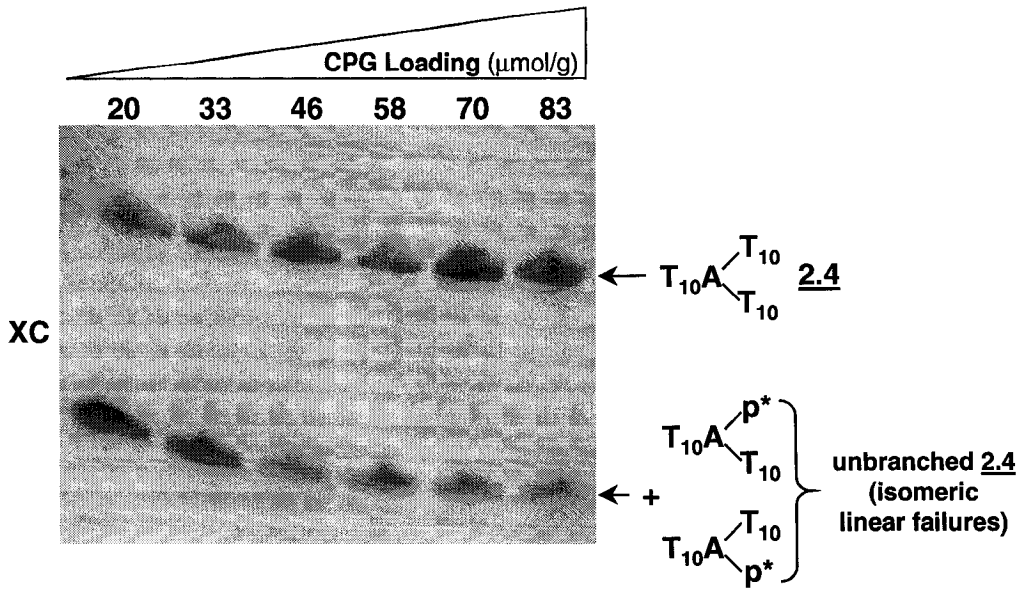


Figure 2.8: Analysis of Q-linked cleavage products by 20% denaturing PAGE (7M urea). Q-linked oligonucleotides were cleaved from the solid-support (CPG) with 50 mM K_2CO_3 in methanol (r.t., 1 min).

under an inert nitrogenous environment. Activation of the 5'-phosphoramidite ensued by the addition of a 0.5 M solution of 1H-tetrazole in CH_3CN, conditions commonly used in the activation of nucleoside phosphoramidites in automated synthesis. Reactions were left shaking for a 16 h period, at which time the CPG was washed exhaustively and dried. Complete deprotection of the bound oligonucleotides was effected using a standard treatment with aqueous ammonium hydroxide solution and the cyclization products partitioned on a denaturing gel (**Figure 2.9**). As seen in **Figure 2.9**, the coupling reaction of the 5'-phosphoramidite resulted in the formation of a number of new product species (compare lanes 2-6 with lane 1). The cyclization of the branched DNA precursors could result in the formation of two regioisomeric lariat species (2',5'-linked and 3',5'-linked lariats) owing to the cleavage of both the 2'-extension or the 3'-extension Q-linkers (**Figure 2.3**; named **2.5 A + B**). Given that the ratio of Q:S linkers on the solid support was nearly 1:1 in all cases, it was expected that the two regioisomers could be formed in equal amounts. Initially, the slowest migrating species were attributed to hyperbranched products, namely products formed by dimerization and catenation reactions between adjacent oligomers. Additionally, two predominant product bands

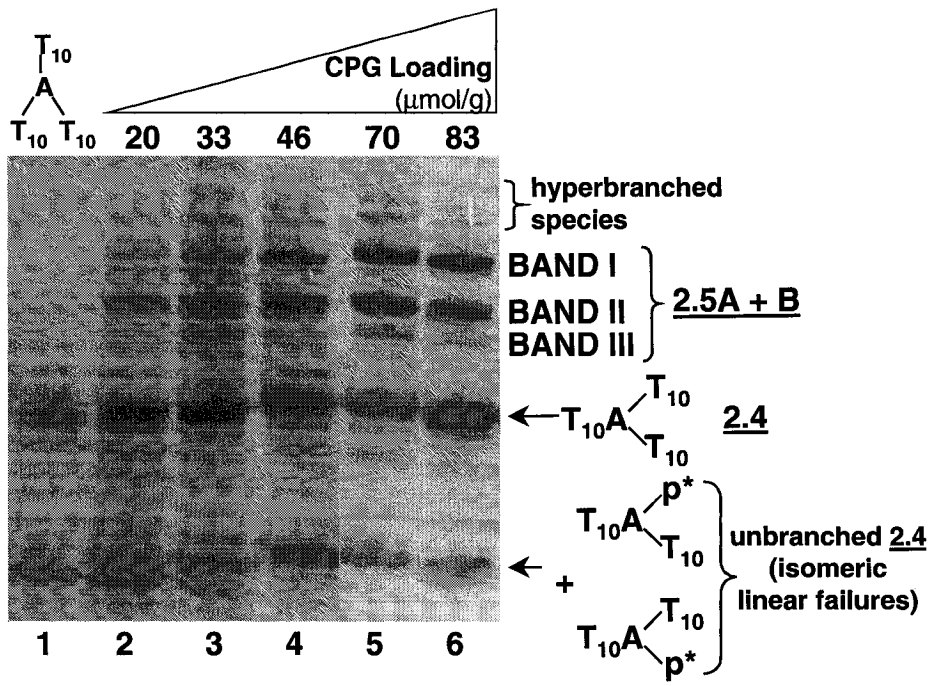


Figure 2.9: 20% denaturing PAGE (7M urea) demonstrating the cyclization of **2.4** on mixed Q/S-linked CPG of various loadings (20-83 μmol/g). Lane 1: crude **2.4** [$T_{10}A^{2',5'}(T_{10})_{3'5'}T_{10}$]; Lanes 2-6: crude cyclization reactions of **2.4** on Q/S-linked CPG of various loadings. The gel bands were visualized by staining with Stains-All® solution.

were clearly evident on the gel (**Figure 2.9**; Bands I and II) when the gel was visualized by UV-shadowing. The retarded electrophoretic mobility displayed by the new products is consistent with reported lariat formation, whereby the circularized structures exhibit frictional hindrance owing to the circularity of the structure.[109] At first, we were excited by this result since it indicted that the two principal reaction products may have corresponded to the 2'- and 3'-linked lariat regioisomers. We were surprised by the possibility that the regioisomeric forms of the lariat could be so effortlessly resolved on the gel since they bore identical molecular weights and sequence composition. Unfortunately, upon staining of the gel, an additional band migrating slightly faster than Band II was also observed (**Figure 2.9**; Band III). Unreacted **2.4** was also evidently present in the reaction mixture, likely co-existing in its 5'-phosphorylated form due to hydrolysis of any unreacted phosphoramidite along with any non-phosphitylated, 5'-

hydroxy starting material. As the CPG loading was variably increased, a noticeable intensification in the amount of Band I and II species was evident (**Figure 2.9**; lanes 2-6). Instead, the amount of Band III product remained unchanged. A credible explanation for this trend will be discussed in the next section.

2.4.3. Characterization of the Cyclization Product Bands

Preparatory gel purification was conducted on one of the cyclization reaction mixtures (83 μmol/g reaction) and the product species corresponding to Bands I-III excised from the gel and desalted by size exclusion chromatography (SEC) on Sephadex G-25®. In an effort to determine which and if any of the product bands corresponded to the desired lariat structure, two distinct enzymatic assays were executed: hydrolysis with a 5'-exonuclease, BSPDE and a 2'-phosphodiesterase, lariat debranching enzyme. Prior to enzymatic characterization, the discrete product species (Bands I-III) were radiolabeled at their free 3'-hydroxy termini with [α-^{32}P]-dideoxyadenosine triphosphate (α-[^{32}P]-ddATP) using the enzyme terminal deoxynucleotidyl transferase (TdT) according to product specifications (**Figure 2.10**). Radiolabeling at the 5'-end of the oligonucleotides was not feasible since the oligonucleotides presumably contained 5'-termini that were tied up in the circularized part of the lariat structure. Attempts were nonetheless made to 5'-end label the molecule with γ-[^{32}P]-ATP and bacteriophage T4 polynucleotide kinase, however, all three product bands were resistant to this reaction, suggesting that either the 5'-termini were inaccessible or a free 5'-hydroxyl unit was not present and instead had been converted to an unrecognizable 5'-moiety (*e.g.* 5'-phosphate; data not shown). Firstly, the circular nature of each of the product species (Bands I-III) was verified by incubating the individual 3'-radiolabeled cyclization products with bovine spleen phosphodiesterase (BSPDE). BSPDE, a 5'-exonuclease, hydrolyzes polynucleotides of both DNA and RNA from the 5'-hydroxyl terminus to produce 3'-monophosphate nucleotides (**Figure 2.11**).[240] Oligonucleotides bearing 5'-phosphate termini are not substrates for the reaction. Lariat structures do not carry any free 5'-termini, therefore, they are not recognized by this phosphodiesterase and are consequently

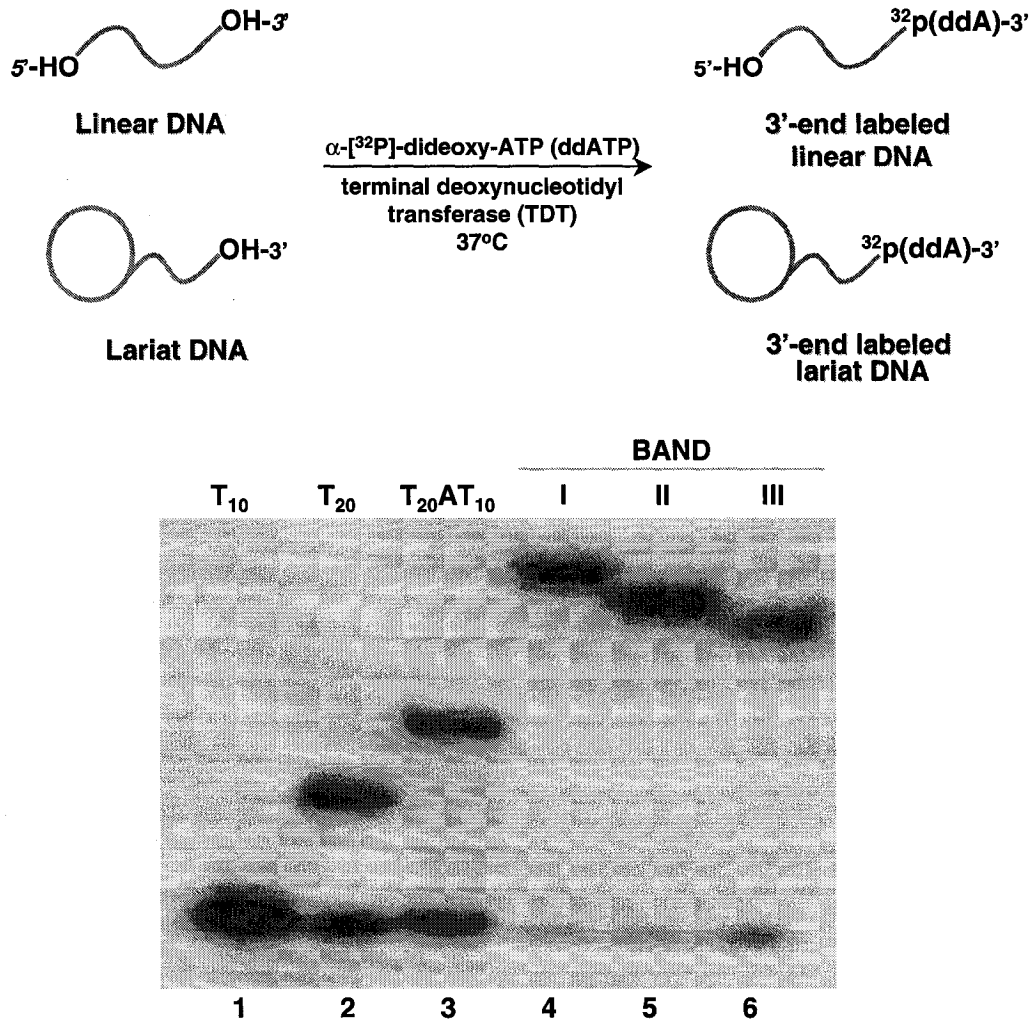


Figure 2.10: Radioactive 3'-end labeling of linear and presumed lariat DNA with α-[^{32}-P]-dideoxy-ATP (ddATP) and terminal deoxynucleotidyl transferase (TDT). Lanes 1 & 2: T_{10} and T_{20} controls; Lane 3: control 2'-phosphodiesterase product (debranching product) of lariat DNA **2.5**; Lanes 4-6: presumed lariat **2.5**, Bands I-III (see **Figure 2.9**)

resistant to enzymatic hydrolysis (**Figure 2.11**). When Bands I-III were incubated directly with BSPDE, no discernable amount of degradation was evident as demonstrated by the lack of any shorter hydrolysis products on a denaturing gel, indicating that the structures conceivably existed as lariats (**Figure 2.11**; lanes 5, 8 and 11). Alternatively, when the Y-shaped bNA **2.4** was treated with the enzyme, the compound was degraded to one main product band suggesting that the enzyme reaches an interruption site at the branchpoint adenosine. This in effect produces a V-shaped DNA molecule, and is attributed to the fact that the enzyme is incapable of processing the oligonucleotide

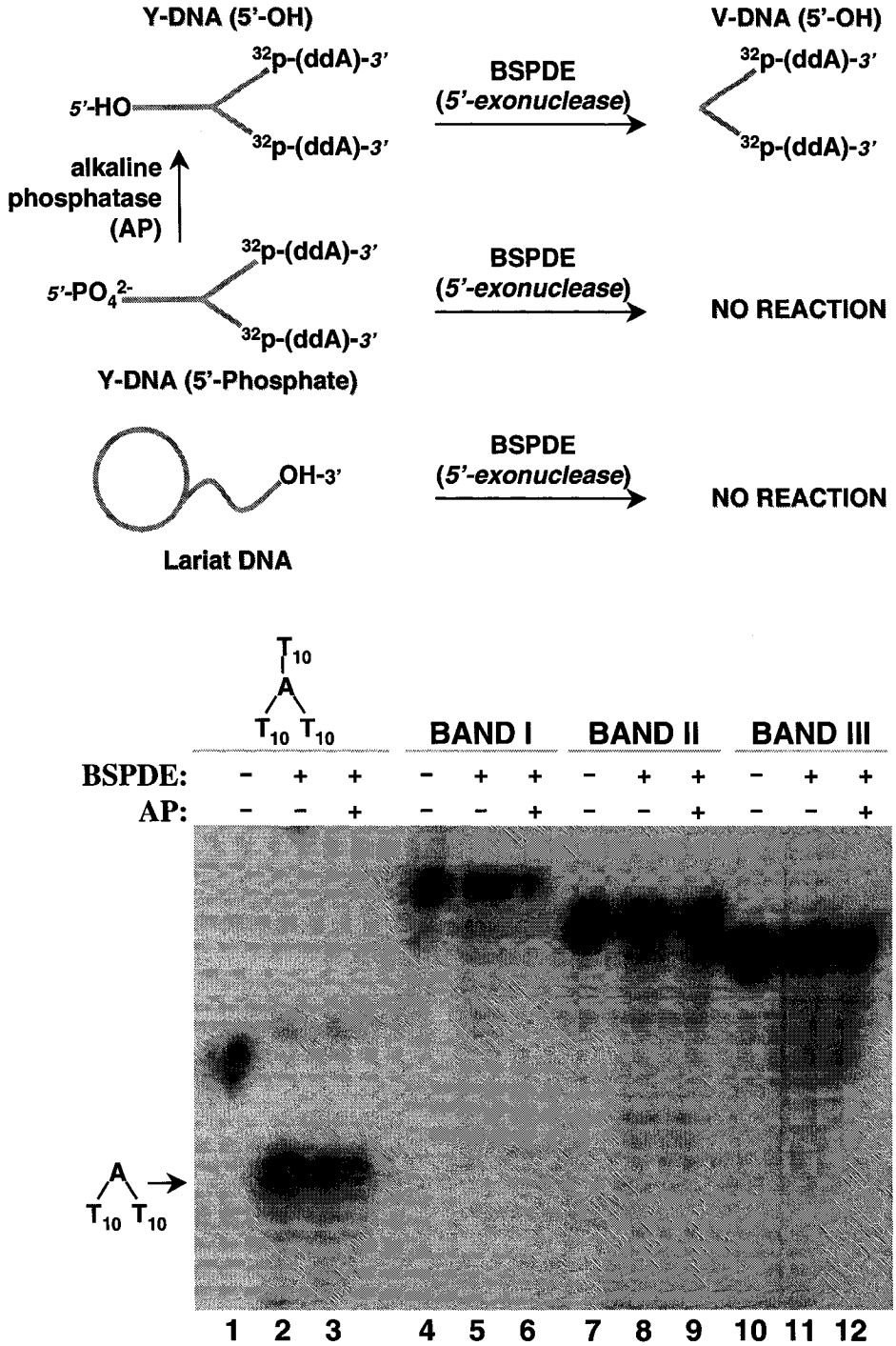


Figure 2.11: Characterization of the branched DNA (Y-DNA) cyclization products by hydrolysis with the 5'-exonuclease, bovine spleen phosphodiesterase (BSPDE). Lanes 1-3: Y-DNA, $T_{10}A^{2',5'}(T_{10})_{3',5'}T_{10}$; Lanes 4-6: cyclization product, Band I (see **Figure 2.9**); Lanes 7-9: cyclization product, Band II (see **Figure 2.9**); Lanes 10-12: cyclization product, Band III (see **Figure 2.9**). Y-DNA is processed from the 5'-terminus to yield a V-shaped DNA product, which is resistant to further exonuclease digestion, owing to the branchpoint. AP= alkaline phosphatase.

further upon reaching the branchpoint.[187,206] Quite possibly, another explanation for the resistance displayed by the three compounds is the fact that they may just be higher-order structures that contain 5'-phosphate groups, and as such, are not substrates for the enzyme. To eliminate this possibility, the three species were incubated with alkaline phosphatase (AP), and enzyme that removes terminal phosphate groups, prior to incubation with BSPDE. The identical result was obtained in all three cases, whereby the species remained resistant to enzymatic hydrolysis (Figure 2.11). Another possible explanation is that the 5'-termini did not include phosphate groups, and instead, hydrogen phosphonates were present. When an activated 5'-phosphoramidite fails to react with the 3'-hydroxyl of the oligonucleotide, trace water becomes the nucleophile, resulting in a hydrogen phosphonate intermediate. However, oxidations of such intermediates to the corresponding phosphate have been shown to occur exceedingly slowly, resulting in predominant H-phosphonate accumulation.[241,242] Since hydrogen phosphonates are not biologically relevant entities, they are not substrates for the enzyme. Seeing as characterization efforts with BSPDE did not produce any conclusive results, an alternative identification means needed to be utilized.

In order to further characterize the nature of the three product species, and to determine if any of the bands corresponded to the DNA lariat, a debranching assay using the specific 2'-phosphodiesterase activity found in HeLa extract, was performed. Specifically, this enzymatic activity, identified in both yeast[113,118] and mammalian[109,112,243] cell lines, converts RNA lariats into linear molecules by exclusive hydrolysis of the 2',5'-phosphodiester linkage, resulting in an oligonucleotide containing a 2'-hydroxyl group and a 5'-terminal phosphate. Unlike other phosphodiesterases (*e.g.* snake venom phosphodiesterase) which cleave both 2',5' and 3',5' phosphodiester bonds, the RNA debranching enzyme (DBR) hydrolyzes only 2',5'-linkages that are vicinal to 3',5'-bonds, thereby leaving all other phosphodiester bonds intact. Since a purified form of the either the yeast[115] or human[112] homologues of the enzyme, was unavailable to us at the time, we exploited the debranching activity found in HeLa nuclear extract (hDBR).[109] EDTA (10 mM) was added to the assay mix as a general exonuclease inhibitor.[109] To

promote more efficient hydrolysis of the 2'-phosphodiester linkage, the cyclization reaction was performed on a Y-DNA similar to **2.4**, however it now incorporated a wild-type G-nucleotide at the 2' and 3'-positions off the branch (**Figure 2.12**). Purine-containing branches have been shown to be the preferred substrates for hydrolysis.[115,116,208,212] The intramolecular condensation reaction produced an identical pattern of product bands on the gel (Bands I-III), which were subsequently 3'-end labeled as described previously. Should any of the three product bands contain one of the regioisomeric lariat products, the individual species could be easily distinguished using this assay since cleavage at the 2'-linkage would result in completely different hydrolysis products (**Figure 2.12**). Specific cleavage at the 2'-phosphodiester bond of the

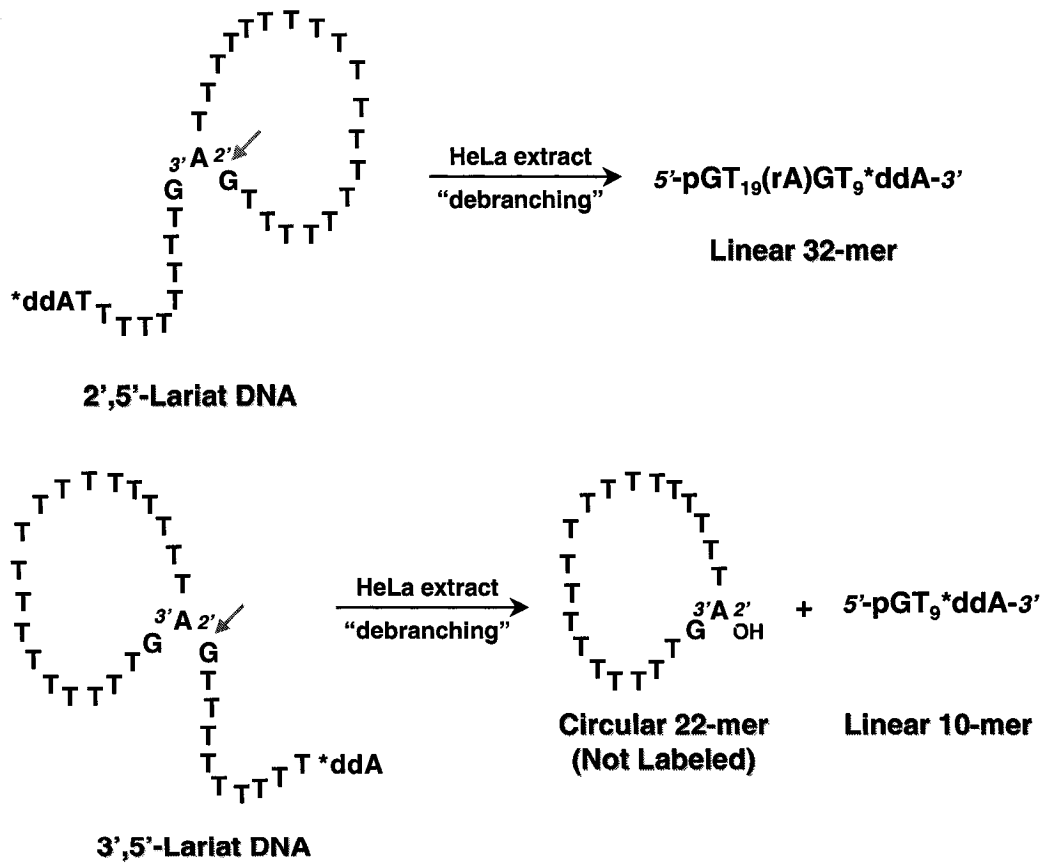


Figure 2.12: Characterization of the 2',5'-linked and 3',5'-linked lariat DNA regioisomers through enzymatic hydrolysis with the 2',5'-phosphodiesterase activity (debranching) present in HeLa nuclear extract. *ddA=3'-radioactive [^{32}P] dideoxyadenosine; p=terminal phosphate. The arrow represents the site of enzymatic hydrolysis.

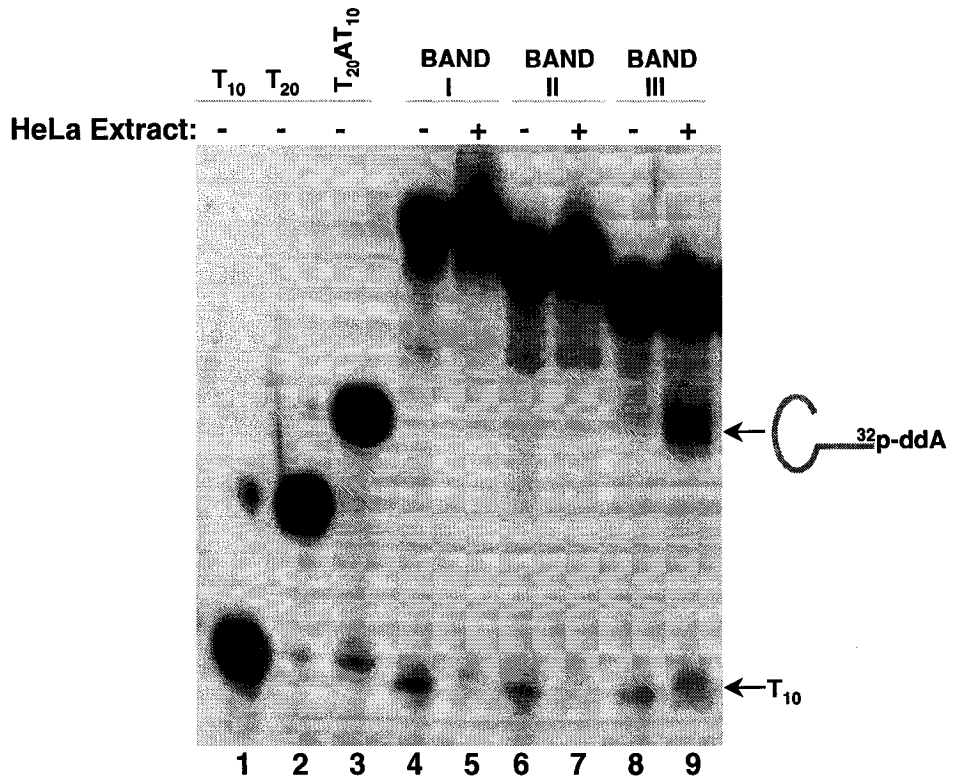


Figure 2.13: Autoradiogram of a 20% denaturing gel (7M urea) demonstrating the debranching of the three isolated cyclization products (Band I, II, and III) with appropriate controls. Lanes 1 & 2: control DNA markers T_{10} and T_{20}; Lane 3: control 2',5'-lariat debranching product marker; Lanes 4, 6 and 8: cyclization products without HeLa nuclear extract (no debranching); Lanes 5, 7 and 9: cyclization products treated with HeLa nuclear extract (debranching).

2'-linked lariat regioisomer would result in the production of a full-length linear 3'-^{32}P radiolabeled construct that is 32-nt long. In contrast, hydrolysis of the 3'-linked lariat would produce two hydrolysis products: a radiolabeled linear decanucleotide and a circular product that does not contain a terminal radiolabel, and as such, would not be visible by autoradiography. The three radiolabeled reaction products (Bands I-III) were incubated with HeLa nuclear extract under debranching conditions[109] and the digestion products resolved by denaturing PAGE (**Figure 2.13**). Bands I and II appeared to be completely impervious to the 2'-phosphodiesterase activity (lanes 5 and 7), however, Band III hydrolysis resulted in the creation of two cleavage products (lane 9). The migratory behaviors of the individual hydrolysis products coincided with both the full-length 2'-lariat debranching product (compare lane 3 and 9) and the 3'-lariat

decanucleotide product (compare lane 1 and 9). Although debranching was not effected to completion, the results indicate that Band III exists as a mixture of the two lariat regioisomers.

Clearly, the debranching assay established that the cyclization product Bands I and II were not consistent with lariat formation. Further examination of the other possible products that could arise during activation and coupling of the support-bound oligonucleotides demonstrated a notable pattern. Presumably, when the Q/S-linker CPG-loading was increased, the distance between adjacent oligonucleotides consequently decreased, intrinsically promoting competing intermolecular reactions between reactive oligonucleotide moieties (**Figure 2.14**). Given that the predominant precursor species in each case consisted of both the fully branched Y-DNA (2.4) and the regioisomeric linear

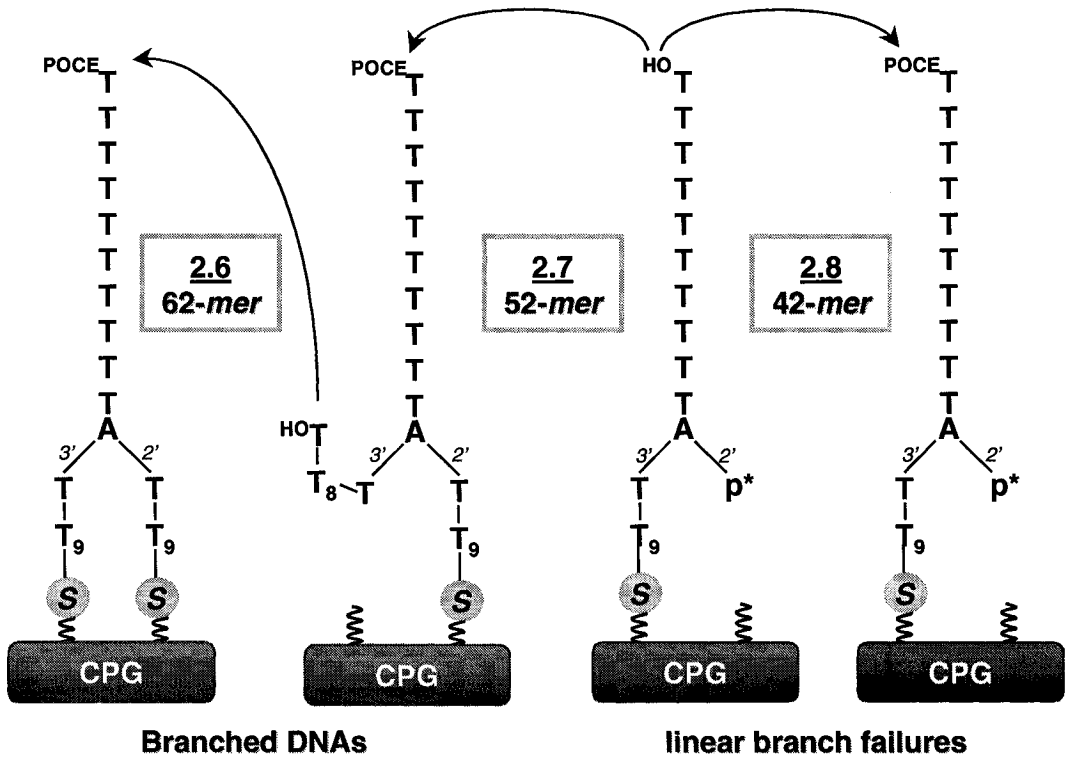


Figure 2.14: Schematic representation depicting the possible dimeric coupling reactions between adjacent support-bound Y-DNA structures and isomeric linear failure sequences. The coupling between two adjacent Y-DNA molecules produces a 62-*nt* hyperbranched species (2.6). Adjacent coupling between a Y-DNA molecule and its corresponding linear isomeric failure sequence yields a 52-*nt* species (2.7). Two linear isomeric failure sequences reacting produces a 42-*nt* oligonucleotide (2.8).

failure sequences of **2.4** (**Figure 2.6**), three main competing intermolecular reactions were possible (**Figure 2.14**). Adjacent coupling between the 3'-hydroxyl of a Q-cleaved Y-DNA and the activated 5'-phosphoramidite of an S,S-linked or Q,S-linked Y-DNA would result in the production of a 62-nucleotide double-branched dendrimer product (**2.6**). Alternatively, reaction amid any unphosphitylated 5'-hydroxyls of the linear isomeric failure sequences, and the 5'-phosphoramidite of a neighboring Y-DNA would produce a 52-nucleotide long oligonucleotide (**2.7**). Similarly, the same isomeric failure sequence could react with an adjoining 5'-phosphitlyated isomeric failure sequence, and give rise to a 42-nucleotide 5',5'-linked extended linear oligonucleotide (**2.8**). Alternative variations to the dendritic products shown are also quite possible, thus complicating matters even further. The resultant dendrimer products boast molecular weights and mass to charge ratios that are significantly higher than the precursor bNA **2.4**, and therefore should exhibit delayed mobilities when resolved on a denaturing gel. Appropriately, two control dendritic oligonucleotides were synthesized in order to elucidate the exact nature of the main products observed after activation and coupling (**Figure 2.15**). Hyperbranched **2.9** was constructed by first synthesizing bNA **2.4** as described previously, and reacting the terminal 5'-hydroxyl of two adjacent bNAs with one equivalent of diluted bis-adenosine phosphoramidite **1.1**. In effect, this "double" branching reaction yields a hyperbranched molecule consisting of 63 nucleotides (**Figure 2.15**).[208] Contrarily, a control 43 nucleotide linear failure branched product was constructed by bridging two linear 21-mers with one equivalent of the bis-adenosine phosphoramidite reagent (**2.10**). In both cases, the crude dendritic reaction mixtures were run alongside the cyclization reaction products (83 μmol/g synthesis) so as to deduce which band corresponded to what sequence according to its migratory behavior on the gel. Visibly, under 20% denaturing gel conditions, the three product bands (Bands I-III) co-migrated with three of the resultant products present in the crude hyperbranched DNA **2.9** synthesis (**Figure 2.16**; compare lanes 3 and 4). Specifically, Bands I and II appeared to directly correspond to compounds **2.6** and **2.7** respectively on a 20% gel. This undoubtedly confirms that as the CPG loading increases, the distance between adjacent oligonucleotides accordingly decreases, and results in predominant dimerization reactions. This increasing dimerization trend was clearly visible on the gel analysis of the cyclization products with

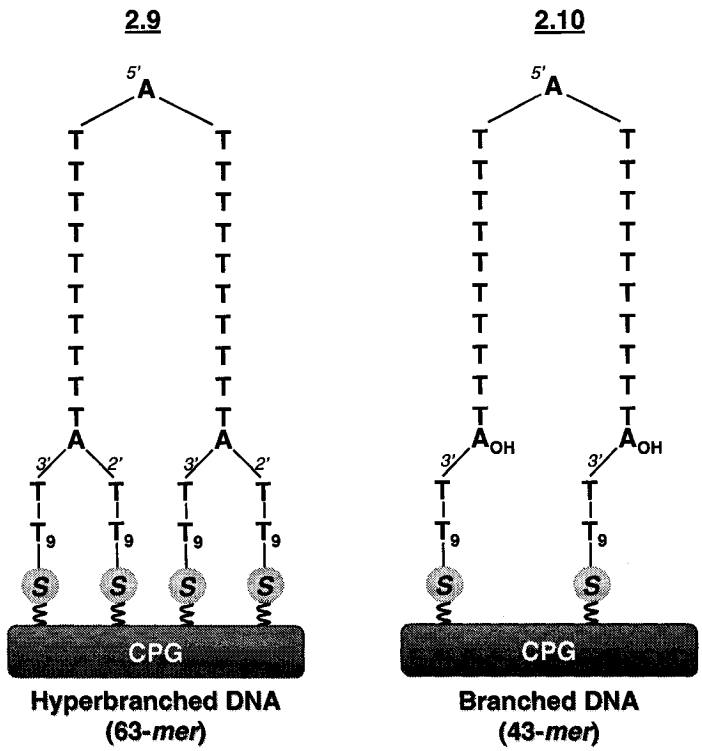


Figure 2.15: Schematic representation of two control dimeric species. On the left, the hyperbranched complex (63-*mer*) is formed by the reaction of two adjacent support-bound Y-DNAs with one equivalent of bis-adenosine phosphoramidite reagent (**2.9**). The 43-*nt* branched (V-shaped) DNA on the right (**2.10**) is formed by the coupling of two adjacent linear DNAs with one equivalent of bis-adenosine phosphoramidite reagent.

escalating CPG-loadings (**Figure 2.9**). A reaction product consistent with the migratory rate of Band III was also evident in the PAGE analysis of **2.10**, indicating that perhaps we were mistaken about the lariat nature of this product. Furthermore, under identical conditions, dendrimeric compound **2.10** and the branched precursor **2.4** were inseparable (**Figure 2.16**; compare lanes 1-4). Plausibly, the two compounds may be resolved on a lower crosslinked gel. Indeed, when a 12% denaturing gel was employed, compound **2.10** exhibited an obstructed migratory behavior compared to bNA **2.4** (compare lanes 1 & 2 and 1 & 3). Most significantly, cyclization product Band III appeared to separate into two distinct bands under the lower crosslinking conditions (**Figure 2.16**; lane 3). One of the bands was clearly manifest in the crude synthesis of the dendritic product **2.9**, whereas the other band did not correspond to any other product in the reaction (**Figure 2.16**; lane 4). In fact, this discrete product band exhibited an anomalous migratory

behavior on diverse crosslinks of native (no urea) and denaturing gels (12-28%), whereas the other intermolecular reaction products (**2.6**-**2.9**) displayed consistent electrophoretic mobility to the control dendritic species **2.9** and **2.10** (not shown). This evidence, as well as the debranching enzymatic assay indicate that under 20% gel conditions, Band III contains two coalesced products: the regioisomeric lariat species and an unidentified product that arises from the intermolecular reaction of two molecules on the solid-support. Under a lower crosslinked gel, the lariat is cleanly resolved from this intermolecular reaction side product.

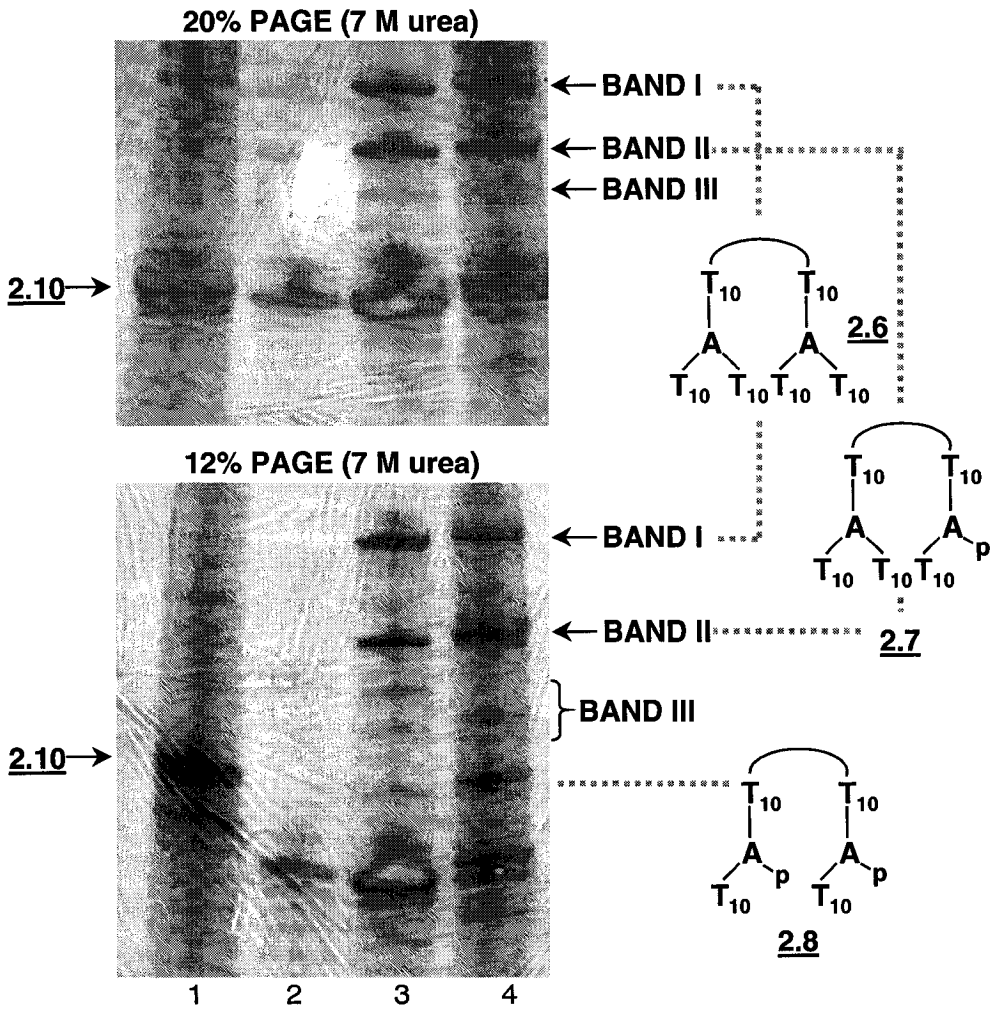


Figure 2.16: Resolution of lariat cyclization bands under different % acrylamide crosslinked gels. Lane 1: control dimeric species **2.10** [$A^{2',5'}(T_{10}rAT_{10})_{3',5'}$ ($T_{10}rAT_{10}$)]; Lane 2: crude **2.4** [$T_{10}A^{2',5'}(T_{10})_{3',5'}$ T_{10}]; Lane 3: crude cylcization reaction of **2.4**; Lane 4: control hyperbranched dimeric species **2.9** [$A^{2',5'}(T_{10}A^{2',5'}(T_{10})_{3',5'}T_{10})_{3',5'}$ ($T_{10}A^{2',5'}(T_{10})_{3',5'}T_{10}$)]. Band III is resolved into two product species under a lower crosslinked (12%) gel.

In addition, MALDI-TOF mass spectrometry was undertaken on each of the products bands, in an effort to confirm the molecular composition of the individual species (**Table 2.1**). Resolution of the individual product bands was conducted on a 12% denaturing

Table 2.1: Experimental molecular weights and retention times of the lariat cyclization product Bands I-III

Compound	Experimental MW (g/mol)	Structure	Length (nts)	Retention Time; t_R (min)[a]
2.4	9475	Y-DNA	31	47
2.5 Band I	18790-18900	Y-DNA Dimer	62	62.5
2.5 Band II	16000	[Y+ Linear] DNA dimer	52	60
2.5 Band III	9455	Lariat DNA	31	47

The cyclization product species Bands I-III were separated and purified on a preparatory denaturing 12% gel (see **Figure 2.16** for migratory pattern). [a]Anion-exchange HPLC conditions: Buffer A: ddH_2O; Buffer B: 1 M $LiClO_4$. Gradient: 10-20% Buffer B over 60 minutes followed by 20-35% Buffer B over an additional 20 minutes. Column temperature: 50°C

PAGE in order to ensure that Band III was well separated from the unidentified dimeric species. The molecular weight of Bands I and II were unmistakably ascribed to intermolecular dendrimer synthesis of either two Y-DNA molecules (**2.6**) or one Y-DNA reacting with a linear isomeric failure sequence (**2.7**). Sure enough, Band III displayed a molecular ion peak that corresponded directly to that of the lariat species **2.5**. Moreover, HPLC analysis of the three product bands also displayed diverse retention behavior on an anion-exchange column. Specifically, the bNA precursor **2.4** and the suspected lariat product **2.5** (Band III) revealed nearly identical retention times. This is not surprising since the difference between the two molecular compositions is merely an extra phosphate charge. Conversely, Bands I and II demonstrated delayed retention times compared to Band III (**Table 2.1**), and were consistent with much larger oligonucleotide structures, and an augmented number of phosphate charges in the dendrimeric backbones. All attempts to separate the regioisomeric 2'- and 3'-lariat species by either PAGE or HPLC were to no avail.

2.5. CONCLUSIONS

The results reported in this chapter clearly indicate that there are several weaknesses with trying to synthesize a medium-sized lariat structure on solid-support using a convergently synthesized branched DNA precursor. Predominantly, the key to a successful branching reaction (*i.e.* high loading CPG) turns out to be detrimental to lariat cyclization and instead promotes the intermolecular linking of adjacent oligonucleotides resulting in principal formation of higher-ordered dendrimer structures. Furthermore, there appears to be a high entropic penalty for randomly cyclizing two decanucleotide strands, even though they are tethered to a solid-matrix.

Nonetheless, we have corroborated that CPG-tethered nucleosides and oligonucleotides can be simply and effectively phosphitylated and used as reactive functionalities for the creation of new phosphodiester linkages (**Figures 2.3** and **2.4**). In addition, we have successfully demonstrated that oligonucleotides bound to a mixed Q and S-linker solid-support system can be selectively cleaved under mild conditions without consequence to the 5'-phosphoramidite or β-cyanoethyl blocked phosphate protecting groups (**Figure 2.8**). As the ratio of Q:S nucleoside linkages present on the support surface was nearly equivalent in all cases, branched DNA synthesis on the mixed resin produces an equal distribution of both Q and S-linked bNA and linear isomeric failure sequences (**Figure 2.6** and **2.8**). Although maximal branching efficiency was attained with high CPG loadings (60-83 μmol/g), this materialized into a significant problem at the activation and coupling stage, whereby predominant dimer formation become evident in lieu of lariat cyclization (**Figure 2.9**). Regardless, a small amount (*ca.* <10%) of lariat product **2.5** was evidently isolated from the complex mixture of reaction products (**Figure 2.16**). Numerous characterization efforts such as a 5'-exonuclease assay, resistance to 5'-end radiolabeling, 2'-debranching reaction, MALDI-TOF-MS, HPLC and its anomalous mobility behavior established that this minor product band displayed key characteristics of the DNA lariat product. In addition, the debranching assay further supported the notion that this band was likely composed of an inseparable mixture of both the 2'- and 3'-lariat regioisomers.

Seemingly, the desired objective for a simplified and operative approach to high yielding lariat oligonucleotide synthesis was not achieved with this method, however, a novel strategy for the synthesis of dendritic-type structures, which bear novel characteristics in their own regard, has been devised. Specifically, high molecular weight dendrimers can be readily assembled on the solid-matrix in one activation and coupling step by simply adjusting the nucleoside CPG loading. Even so, it became increasingly clear that in order to successfully synthesize lariat oligonucleotides, an approach that minimizes the entropic penalty of cyclization and aligns the reactive termini for effective fixation must be adopted. This successful methodology will be discussed in the next chapter.

CHAPTER 3: TEMPLATE-MEDIATED CHEMICAL SYNTHESIS OF CIRCULAR AND LARIAT OLIGONUCLEOTIDES

3.1. INTRODUCTION

Evidently, there were two major shortcomings with the intramolecular methodology described in Chapter 2 for the circularization of a DNA lariat on solid-support. Firstly, since oligonucleotide synthesis was conducted on a solid-matrix of defined functional group spacing, the oligonucleotide strands to be cyclized were in close proximity to adjacent support-bound polymers, and as such, intermolecular reaction prevailed. Furthermore, the untethered decanucleotide strands are freely rotating entities, such that intramolecular association is an entropically disfavored process. Given these limitations, we looked towards cyclization approaches that would still exploit the topology of branched oligonucleotides as lariat precursors, however utilizing methods that would force the reactive termini of the 2' and 5' appendages off the branch in a proximal orientation for ligation.

A number of research efforts have been devoted to methods for synthesizing circular oligonucleotides. For practical purposes, chemists have been primarily interested in the structure and reactivity of circular DNA and RNA oligonucleotides smaller than those which occur in nature (*e.g.* supercoiled DNA, RNA viroids).[1,244] Duplex DNA can exist in circles as small as 125 base pairs[245], however realization of cyclic structures smaller than this is difficult owing to the rigidity of the double helix. In contrast, single-stranded circular DNAs and RNAs as small as two nucleotides have been reported owing to their innately more flexible structure.[246] The advent of solid-phase DNA and RNA synthetic methodologies has made the sphere of possible cyclic DNA and RNA structures fall into a size range most attractive to chemists. A significant number of reports have demonstrated how circular oligonucleotides can have properties very different from their linear correlatives. Among these are enhanced DNA-binding affinity[247-249], greater sequence selectivity[248,250,251], more pronounced resistance to degradation by

ubiquitous cellular nucleases[252], and the ability to serve as competent templates for DNA[253] and RNA[254] polymerases in rolling circle amplification. From a therapeutic perspective, cyclic oligonucleotides have been used as sequence-specific transcription factor decoys[255-257], as well as antisense regulators of gene expression by binding to their single-stranded targets *via* either duplex[258-260] or triplex[247,249,251 261] forming abilities.

As the cyclization reaction from a synthetic linear precursor molecule is key to producing appreciable amounts of circular oligonucleotide for physical and biological studies, numerous groups have investigated diverse methods for optimizing the phosphodiester bond forming reaction by either enzymatic[262,263] or chemical ligation methods.[232,264-268] Many of these random cyclizations (non-templated) suffer from only modest yields, as a result of the high entropic penalty of forming rings larger than 10 nucleotides. Furthermore, high dilution conditions or the tethering of the precursor to a solid-support were requisite in order to prevent competing intermolecular amalgamation. Conversely, approaches involving the use of oligonucleotide templates have demonstrated considerable success for forming larger ring structures by either chemical or enzymatic ligation means.[249,258,259,269-273] Fundamentally, a single stranded linear oligonucleotide (precursor circle) is designed to be effectively recognized by a complementary template strand (splint) such that it forms a stable hybrid complex (**Figure 3.1**). When the precursor is an oligonucleotide consisting of both a free phosphate and hydroxyl terminus, association renders the reactive moieties in a flanking alignment for efficient ligation. Such templated reactions are highly constructive owing to the favorable enthalpic contributions from hydrogen bonding and base-stacking in the complex, which compensates for the unfavorable entropy of bringing two oligonucleotide strands together, thereby positioning the reactive phosphate and hydroxyl termini into close proximity.[274] Both double[258,271] and triple-helical[247-249,273,275] complexes have been used to assemble circular oligonucleotides, however in the latter case, their synthesis is limited to those precursors that contain homogenous pyrimidine or purine tracts. As such, the sequence differentiation within the circle itself is limited. More

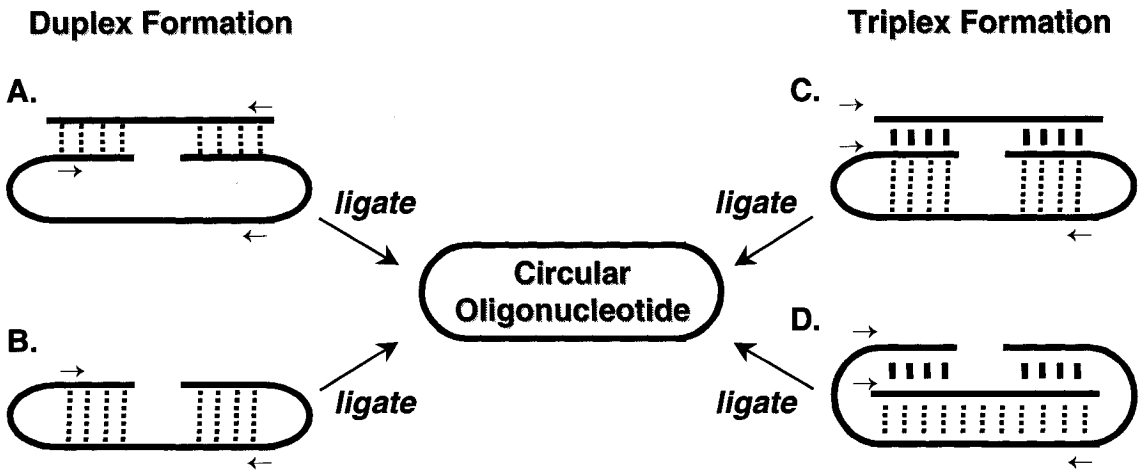


Figure 3.1: Template mediated ligation of circular oligonucleotides *via* duplex or triplex complex formation. Watson-Crick base pairs are denoted by hashed lines. Hoogsteen complementarity is denoted by solid lines. Arrows represent the strand orientation, either parallel or anti-parallel. Panel A: bimolecular "splint" mediated circularization; Panel B: Unimolecular dumbbell complex circularization; Panel C: Triple helical complex circularization whereby the splint oligonucleotide binds *via* Hoogsteen complementarity only; Panel D: Triple helical complex circularization whereby the splint oligonucleotide binds the precursor via both Watson-Crick and Hoogsteen base-pairs.

recently, a circularization effort based on the DNA *i*-motif, a four-stranded oligonucleotide assembly, has been shown to direct the sequence-specific formation of a phosphodiester linkage thereby epitomizing a new type of structural template for constructing circular oligonucleotides.[276,277] Although high yielding circularization has been reported, the sequence design is limited to those oligonucleotides bearing long tracts of cytosine residues. In an alternative fashion, the templates may be included within the single strand itself (*i.e.* self-templating), resulting in the formation of a dumbbell-shaped (*i.e.* double-hairpin) complex, which are internally base-paired and closed off by two stabilizing hairpin loops (**Figure 3.1B**).[278] To date, DNA dumbbells have been synthesized enzymatically using two diverse strategies; ligation of two sticky ended hairpin molecules[279,280], or intramolecular ligation of a nicked dumbbell complex.[281,282] Both ligation methods entail the use of T4 DNA ligase; an ATP-dependent enzyme that covalently joins nicks in double-stranded DNA, and requires a 5'-phosphoryl group for ligation to a 3'-hydroxyl. As such, ligation is contingent upon the

size of the dumbbell to be ligated, the sequences in the hybrid portion, and is limited by the expense of DNA ligase, making enzymatic procedures undesirable for large-scale production. More effectively, nicked junction ligation with purely chemical condensing reagents (*e.g.* EDC, cyanogen bromide) conveniently and inexpensively produces circularized dumbbell without the many limitations inherent to enzymatic procedures.[269,270,283]

Given this wealth of information on the circularization properties of oligonucleotides, we imagined that such template approaches could be exploited in the synthesis of DNA and RNA lariat oligonucleotides. Specifically in this chapter, we devoted our attention to synthesizing such intronic intermediates using either a self-assembled dumbbell complex (**Section 3.2**) or by way of a bimolecular templated reaction (**Section 3.3**). The starting precursor in all cases was a branched DNA or RNA oligonucleotide containing both a phosphorylated and hydroxylated terminus in order to effect a new phosphodiester bond, thereby cyclizing the structure. The design is highly favored over the non-templated method described in Chapter 2 as it auspiciously places the reactive termini into an appropriate spatial geometry for effective ligation, thereby minimizing the number of intermolecular aggregate side products. Furthermore, in the case of the bimolecular template approach, the flexible sequence design allows us to incorporate virtually any nucleotide within the stem and loop portions of the lariat, thereby permitting us to synthesize lariats of any sequence composition as well as the more biologically relevant RNA lariat.

3.2. Synthesis of Lariat DNA via the Chemical Ligation of a Dumbbell Complex

3.2.1. Experimental Design and Project Objective

In order to synthesize a DNA-lariat based on an intramolecular template approach, we looked towards the hybridization characteristics of DNA dumbbells of the type shown in **Figure 3.2**. The dumbbells of interest consist of single stranded DNA that is either phosphorylated at the 3' or the 5'-end, and contains regions of self-

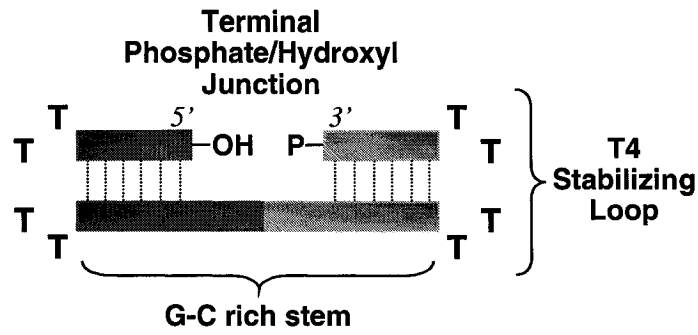


Figure 3.2: Schematic representation of the intrinsic structural features of DNA dumbbells synthesized in this study.

complementarity along the single-strand. Preferably, these regions are G-C rich and as such, can form more thermodynamically stable duplex structures. Additionally, further stabilization is imparted by the incorporation of a T_4 closing loop. This tetrameric loop-structure, along with its pentameric (T_5) counterpart, has been shown to be the most thermodynamically stabilizing loop structures of all the simple, homonucleotide-containing loops.[280,284,285] It has been assessed that the two terminal thymidines of the tetranucleotide loop are involved in a T-T wobble base pair.[285] Upon association of the dumbbell structure, the terminal phosphate and hydroxyl group are proximally aligned for intramolecular condensation (**Figure 3.3A**) by either chemical[269,270] or enzymatic means (T4 DNA or RNA ligase).[282,286] Ideally, the necessary functionality (*i.e.* phosphate introduction) and phosphodiester bond formation would be conducted chemically, so as to reduce cost and allow for an increased yield of product.

Linear dumbbell sequences were designed to favor exclusive intramolecular duplex formation. The preferred dumbbell secondary structure of the designed sequences was confirmed using the DNA MFOLD server, a program that calculates the estimated ΔG for duplex formation in a DNA sequence containing self-complementarity.[287,288] (available at: http://www.bioinfo.rpi.edu/applications/mfold/old/dna/form1.cgi) The initial dumbbell sequence (**3.4**-**3.7**) was selected so that direct comparisons could be made with a similar complex previously studied by Ashley and Kushlan.[269] Incorporated within the

this dumbbell sequence is a 2',5'-linked riboadenosine (and complementary thymidine) which mimics the conserved residue found at the branchpoint in naturally occurring lariat structures. The linear dumbbells were studied prior to lariat synthesis with the aim of evaluating the following: (1) the effect of incorporating a 2',5'-riboadenosine (2'-rA) insert and dT_4 hairpin loop on the thermal stability of the complex; (2) the ligation efficiency of a 3'-phosphate/5'-OH *versus* a 5'-phosphate/3'-OH; (3) the optimal nucleotide composition at the ligation junction, and (4) the effectiveness of ligation directly across from or distal to the 2',5'-rA insert.

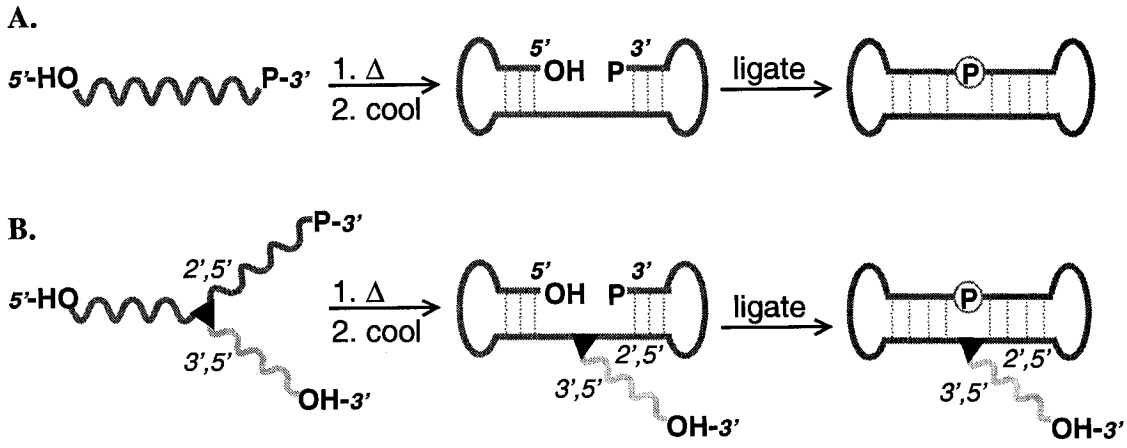


Figure 3.3: Association and proximal alignment of the phosphate/hydroxyl junction in (A) linear and (B) branched DNA dumbbells for the chemical ligation of circular and lariat DNA.

The approach adopted for lariat DNA synthesis used commercially available phosphoramidite building blocks in conjunction with our previously published protocol for the controlled regiospecific solid-phase synthesis of branched DNA molecules[212,221,233], phosphitylation of the 3'-terminus to yield a 3'-phosphate and chemical condensation of the nicked phosphate/hydroxyl junction. The lariat precursor is designed such that the 5' and 2' extensions of the branched (Y-shaped) precursor are capable of intramolecularly folding into a "dumbbell" complex, thereby aligning the reactive 5'-OH and 3'-phosphate for chemical ligation (**Figure 3.3B**). The 3' extension of the molecule was constructed so that it did not share sequence complementarity with

any part of the 5' or 2' extensions, and as such, is excluded from the dumbbell complex. This allows us to circumvent the formation of the 3',5'-linked lariat regioisomer.

3.2.2. Synthesis of DNA Dumbbell Precursors

Linear DNA dumbbell precursors were synthesized on an Applied Biosystems 381A oligonucleotide synthesizer using standard phosphoramidite chemistry. Primarily, the sequences were comprised of deoxynucleotide residues with the integration of a single riboadenosine (rA) unit, linked by either a 3',5'- or a 2',5'-phosphodiester bond to its DNA neighbour (**Table 3.1**:). The 3'-rA component was introduced during DNA synthesis with standard 5'-O-dimethoxytrityl-2'-O-*tert*-butyldimethylsilyl-(N^6-benzoyl adenosine)-3'-O-(β-cyanoethyl) N,N'-diisopropyl phosphoramidite while the corresponding 2'-rA substituent utilized the 2'-O-phosphoramidite analogue (**Figure 3.4**). It has previously been shown by our group[289,290] and others[291] that the integration of 2',5'-RNA constituents into an otherwise unmodified oligonucleotide duplex results in a decrease in the thermal stability (T_m) of the complex. Rationally, both the 3',5'- and the 2',5'-rA substitutions were studied in order to assess the effect of including the destabilizing 2'-rA surrogate into the overall complex stability. The conditions for coupling and the concentrations of rA-phosphoramidites were identical in both cases, however, the coupling time of the 2'-rA phosphoramidite was extended to 15 min compared to the 10 min required for the 3'-rA moiety. Typically, 1-H tetrazole (pKa 4.9) is used in the activation of nucleoside 3'-phosphoramidites, however, it has been shown that coupling is exceedingly poor for 2'-phosphoramidites using this reagent.[290] Alternatively, a more acidic activator, 5-ethylthio-1H-tetrazole (pK_a 4.28)[292] has been used in the place of 1-H tetrazole in the synthesis of 2',5'-RNA as it has been shown to accelerate protonation of the diisopropylamino group of the phosphoramidite thereby producing a highly electrophilic intermediate that can react with a nucleophilic 5'-hydroxyl group.[293,294] Reportedly, increased coupling yields for 2'-O-phosphoramidites are possible using this reagent compared with tetrazole (98% *vs* 90% average coupling yields), however, the increased acidity of the reagent results in the partial detritylation of the phosphoramidite monomer during synthesis, and the production

Table 3.1: Sequences and extinction co-efficients of linear and branched DNA dumbbells

Code	Sequence (5'→ 3')	$\varepsilon \times 10^4$ (mol/L·cm)
Bimolecular Complexes		
3.1	^{HO}cgc-$A_{3'5'}$-cgc$_{OH}$	6.04
3.2	^{HO}cgc-$A^{2'5'}$-cgc$_{OH}$	6.04
3.3	gcgtgcg	6.37
Unimolecular (Dumbbell) Complexes		
3.4	^{HO}gcg-t_4-cgc-$A_{3'5'}$-cgc-t_4-gcgt$_{OH}$	18.91
3.5	^{HO}gcg-t_4-cgc-$A^{2'5'}$-cgc-t_4-gcgt$_{OH}$	18.91
3.6	^{P}gcg-t_4-cgc-$A^{2'5'}$-cgc-t_4-gcgt$_{OH}$	18.91
3.7	^{HO}gcg-t_4-cgc-$A^{2'5'}$-cgc-t_4-gcgt$_P$	18.91
3.8	^{HO}tgcg-t_4-cgca-$A^{2'5'}$-cgc-t_4-gcgt$_P$	20.86
3.9	^{HO}tgttgcg-t_4-cgcaaca-$A^{2'5'}$-cgc-t_4-gcgt$_P$	26.71
3.10	^{HO}tgcg-t_4-cgcaaca-$A^{2'5'}$-cgc-t_4-gcgttgt$_P$	26.71
Unimolecular (Dumbbell) Lariat Complexes		
3.11	^{HO}gcg-t_4-cgc-$A^{2'5'}$-(cgc-t_4-gcgt$_P$)$_{3'5'}$-t_{12}	28.68
3.12	^{HO}tgcg-t_4-cgcaaca-$A^{2'5'}$-(cgc-t_4-gcgttgt$_P$)$_{3'5'}$-t_{16}	39.77

Notation: small caps=deoxynucleotide residues; $A^{2'5'}$=2,5'-linked ribo-adenosine; $A_{3'5'}$=3'5'-linked ribo-adenosine; P=terminal phosphate; OH=terminal hydroxyl. Bracketed sequences represent those strands linked by a 2',5'-phosphodiester bond to the branchpoint adenosine unit. Extinction coefficients (ε) were obtained from a biopolymer calculator (http://paris.chem.yale.edu/extinct.html) and utilized the nearest-neighbor approximation. ε values for 2',5'-linked units were assumed to be the same as 3',5'-linked units. ε values for branched oligonucleotides were assumed to be identical to those for their full linear constructs.

of highly undesired "n+1" sequences or "long-mers".[290] In an effort to develop a more nucleophilic rather than acidic activating reagent for oligonucleotide synthesis, Vargeese and co-workers at NeXstar Technology Products (Boulder, CO) identified 4,5-dicyanoimidazole (DCI) as an effective and versatile replacement to the commonly used tetrazole.[295] DCI is less acidic (pK_a 5.2) yet more nucleophilic than both 1-H-tetrazole

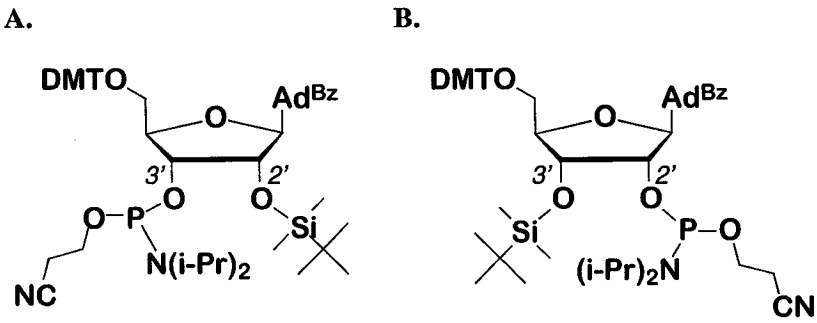


Figure 3.4: Structure of the (A) 3'-riboadenosine phosphoramidite and (B) 2'-riboadenosine phosphoramidite used in the synthesis of circular and lariat DNA dumbbells.

and 5-ethylthio-1-H-tetrazole, and in turn accelerates the rate-determining displacement of the diisopropylammonium group from the activated nucleoside phosphoramidite. In fact, a two-fold increase in activation and coupling of nucleoside phosphoramidites has been reported when DCI was used in the place of tetrazole.[295] Appropriately, DCI was used in the synthesis of the linear DNA dumbbell chimeras and resulted in the production of one predominant product band, which was easily isolable from any shorter failure sequences (**Figure 3.5**).

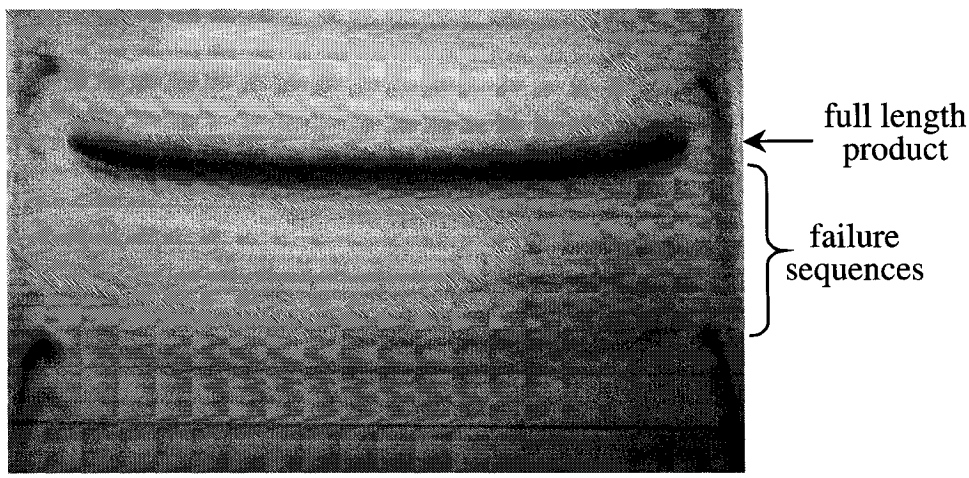


Figure 3.5: 20% denaturing (8.3 M urea) page analysis demonstrating the effective synthesis DNA dumbbell precursor **3.8** with 4,5-dicyanoimidazole (DCI).

Terminal phosphate groups (either 3'- or 5'-phosphates) were introduced chemically into the nicked dumbbell precursors using the phosphorylation reagent, 2-[2-(4,4'-dimethoxytrityloxy)ethylsulfonyl]ethyl-(2-cyanoethyl)-(N,N-diisopropyl)-phosphoramidite (*i.e.* 5'-Phosphate-ON reagent).[296] This commercially available reagent allows for the efficient and reliable introduction of a phosphate group at the 5'-terminus of a support-bound oligonucleotide chain. Furthermore, it is a time- and cost-efficient alternative to enzymatic 5'-phosphorylation of a purified oligonucleotide sample with ATP and the phosphoryl transfer enzyme, T4 polynucleotide kinase. When a 5'-phosphate was desired (**3.6**) the phosphorylation reagent was coupled to the free 5'-hydroxyl of the last nucleotide in the sequence (**Figure 3.6A**). The concentration and

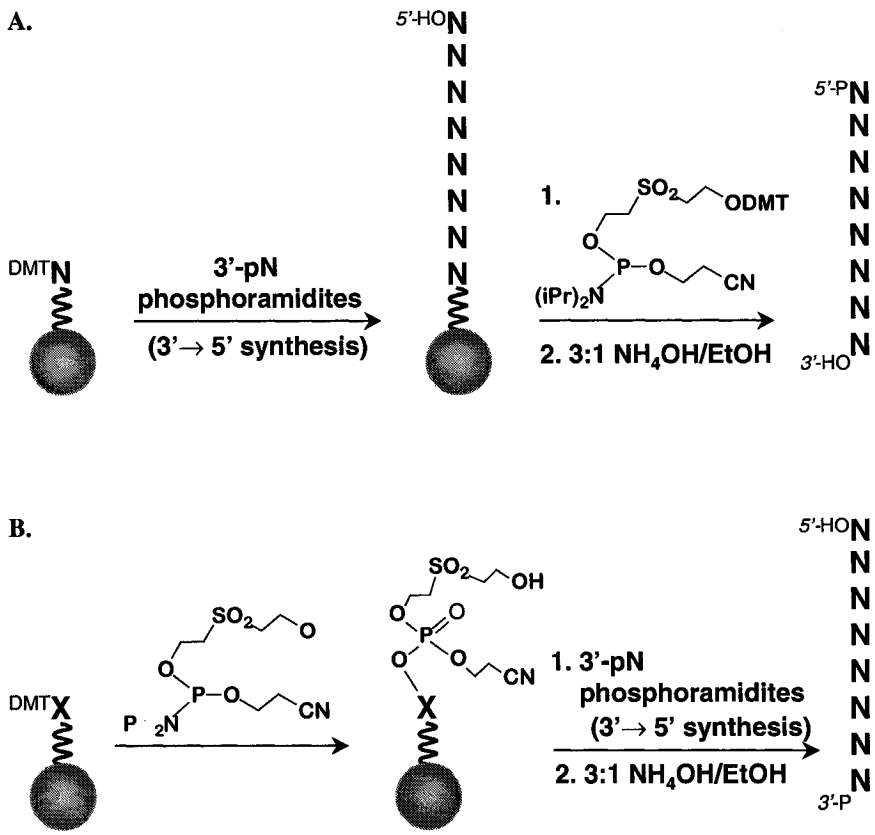


Figure 3.6: Schematic representation of the synthesis of terminally phosphorylated oligonucleotides using the 5'-Phosphate-ON reagent (Chemgenes Corp.). Panel A: Synthesis of 5'-phosphorylated oligonucleotides. Panel B: Synthesis of 3'-phosphorylated oligonucleotides. N=nucleotides present in the desired sequence; 3'-pN=3'-nucleoside phosphoramidites (DNA or RNA); X=any nucleoside; OH=terminal hydroxyl; P=terminal phosphate.

coupling times used were identical to those used for DNA phosphoramidites (0.1 M, 90 s coupling). Since the phosphorylation reagent bears a dimethoxytrityl group, it is possible to monitor its coupling efficiency, which is typically very high (97-99%). Treatment of the 5'-phosphite triester under standard deprotection conditions (3:1 NH_4OH/EtOH) results in the β-elimination of the ethyl-sulfonylethyl phosphate protecting group[297], and concurrent abstraction of the β-cyanoethyl phosphate protecting group. Consequently, a 5'-phosphate monoester is produced.

Alternatively, introduction of a 3'-phosphate (**3.7-3.10**) required that the phosphorylation reagent be coupled firstly to a support-bound nucleoside followed by chain-extension of the desired sequence from the free hydroxyl of the sulfonylethoxy group (**Figure 3.6B**). Following ammonia deprotection, the oligonucleotide includes a 3'-phosphate group attached to the second base added during the synthesis. Hence, the type of nucleoside bound to support is of no consequence, since deprotection results in cleavage of both the support-bound nucleoside and the chemical phosphorylation reagent itself.

Cleavage of the oligonucleotides from the support, removal of exocyclic amino and phosphate protecting groups was conducted under standard deprotection conditions as mentioned previously. However, for those molecules containing 3'-rA or 2'-rA residues, an ensuing treatment with the desilylation reagent, triethylammonium hydrofluoride (TREAT-HF) was required to remove the 2'- or 3'-TBDMS protecting group. Sequences were analyzed and purified on denaturing polyacrylamide gels (20%) that contained a high concentration of urea (8.3 M) as a denaturant, such that the oligonucleotides were resolved in their completely random-coiled state (Experimental section 7.5.2.). Gel salts and urea were removed by size-exclusion chromatography on Sephadex G-25® (Experimental section 7.5.5.1.) prior to characterization of the nicked dumbbells by MALDI-TOF mass spectrometry (Experimental section 7.5.6).

3.2.3. Effect of the T_4 Loop and 2,5'-Riboadenosine Substitution on Dumbbell Stability

The effect of incorporating a single 2'-riboadenosine unit into a self-complementary DNA oligonucleotide was evaluated using thermal melting analysis. As an oligonucleotide complex is slowly heated at a constant UV-wavelength (λ=260 nm), the transition between an ordered native structure and a disordered, denatured state can be monitored.[14] The visible increase in absorbance is indicative of the unstacking and disruption of the base pairs along the helix axis (*i.e.* hyperchromicity) and specifies the relative orderliness of the secondary structure. The midpoint of the sigmoidal transition is defined as the melting temperature (T_m) of the complex, and indicates the point at which half of the oligonucleotide population is in its random-coiled state. Initially, the destabilizing effect of substituting a 2'-rA unit (**3.2**) for a 3'-rA unit (**3.1**) in a set of linear heptanucleotides was monitored by first hybridizing each of the strands with their DNA complement **3.3** (**Table 3.2**). These linear duplex sequences mirrored the stem parts of the ligated dumbbells **3.4** and **3.5**. Not surprisingly, such bimolecular associations resulted in a 5°C destabilization of duplex **3.2**:**3.3** over **3.1**:**3.3**. Previously, it had been demonstrated that one insert of 2'-rA into an rA_{10} homopolymer resulted in a decrease of –4.2°C/insert when bound to poly-dT.[290] Although 2',5'-RNA nucleotides appear to favor the same C-2' *endo* sugar pucker as 3',5'-DNA[298,299], they demonstrate reduced affinity for deoxynucleotide units in a complementary strand than their corresponding 3',5'-RNA correlatives.[289,290,298] While the 2',5'-RNA and 3',5'-DNA boast the same sugar pucker, a key structural distinction lies in the C2'-C3' bond extension, which places the 2',5'-linkage 7 bonds away from its adjacent 5'-phosphate (compared to 6 bonds in a 3',5'-linkage). This additional bond provokes a more "compact" structure and decrease in the magnitude of intranucleotide P-P distance (P-P= 5.9 Å)[298] for the 2'5'-rA compared to the 3',5'-DNA moieties (P-P=6.7 Å)[2], thereby introducing a point of conformational rigidity and local distortion of the helix axis. On the other hand, when we compare nicked dumbbell **3.4** to its homologous counterpart, **3.5**, which includes one 2'-rA insert, the drop in T_m was 3°C, and the T_m's for both complexes were at approximately 20°C higher than their corresponding bimolecular complexes which lacked the T_4 loop.

Table 3.2: Summary of T_m Values and Yields for DNA Dumbbell Complexes

DNA Chimera	T_m of DNA Complex (°C)		Nucleotides at Nicked Junction	% Yield Ligated Product[a]
	A nicked	B ligated		
3.1[b]	35			
3.2[b]	30			
3.4	53			
3.5	50			
3.6	50	>85	GpT	4[c]
3.7	n.m.	>85	GpT	10[c]
3.8	54	87	TpT	66[d]
3.9	62	88	TpT	79[d]
3.10	54	88	TpT	88[d]
3.11	n.m.	n.m.	GpT	n.m.
3.12	52	86	TpT	37[e]

T_m's were obtained in 10 mM Tris-HCl pH 7.5, 10 mM NaCl buffer. Temperatures are the averages of two successive runs and are within ±1°C. [a]The % yield is the ratio of ligated dumbbell to unreacted precursor. [b]The T_m values of **3.1** and **3.2** were obtained by hybridization to their corresponding DNA complement **3.3**. [c]Compounds **3.6** and **3.7**: Ligations were performed with N-cyanoimidazole and the amount of ligated product determined by densitometry (UN-SCAN-IT Software, Silk Scientific). [d]Compounds **3.8**-**3.10**: Ligations were performed with CNBr and the extent of ligation determined by HPLC integration of the product peak. [e]Compound **3.12**: Ligation was performed with CNBr and the amount of ligated product determined by densitometry. n.m.=not measurable.

This implies that the homonucleotide loop cooperatively stabilizes the dumbbell structures thereby sustaining the 2'-rA substitution to a greater extent than the linear duplex analogues. The resultant T_m of **3.5** (50°C) established that the 2'-rA insertion was well tolerated in the dumbbell state, and allowed us to proceed with investigating its ligation behavior.

3.2.4. Optimization of DNA Dumbbell Ligation

A key problem in the template-mediated chemical ligation of oligonucleotide substrates has concerned the development of phosphate monoester activation methods in aqueous media. Internucleotide linkage formation requires a succession of two reactions:

activation of the phosphate group followed by condensation of this activated phosphate with the free hydroxyl group of a nucleoside or nucleotide. In aqueous solutions, it is anticipated that a competing reaction with water for the activated phosphate would prevent the condensation reaction from occurring. As such, formation of a phosphodiester bond would only be expected if the activated phosphate and hydroxyl groups were in a proximal arrangement. Template-mediated alignment of the nicked phosphate-hydroxyl junction allows one to hold the two reactive moieties in a favorable, adjacent position for activation and erection of a new phosphodiester bond.

In the early 1980's, Shabarova and co-workers described the use of the water soluble 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), a commonly used peptide coupling reagent, as an effective mediator of template-directed condensation of proximal phosphate-hydroxyl junction at a nick site.[300-302] Although high yielding phosphodiester bond construction has been reported, its application suffers from a slow rate of reaction (0.5-6 days)[303] and the significant danger of heterocyclic base modification of single-stranded regions in nucleic acids.[300,304] A few years later, two independent research groups were investigating the use of cyanogen bromide (CNBr) as a possible phosphodiester condensation reagent in aqueous solutions.[305,306] Cyanogen bromide, a derivative of hydrogen cyanide (HCN), had previously been suggested as a possible condensing reagent by Orgel and Lohrmann in the prebiotic synthesis of RNA oligonucleotides.[307] In both studies, imidazole was a necessary component of the reaction mixture since reactions that lacked the co-reagent produced undetectable amounts of oligonucleotide. It was clear in both studies that a new compound, N-cyanoimidazole (CNIm) was being formed from the reaction between imidazole and CNBr, and that it was the likely candidate involved in the phosphate-activation mechanism. When purified forms of N-cyanoimidazole were used, the yields observed were appreciably higher than those observed with mixtures of CNBr and imidazole, however long reactions times were still requisite.[308] At approximately the same time, Shabarova and co-workers devised of CNBr-mediated ligation conditions which preclude the need for imidazole as a co-reagent, and were able to produce new phosphodiester

linkages in the order of minutes rather than hours.[309] Remarkably, the efficiency of nick ligation in a DNA duplex was 96% using a 10^3-fold excess of CNBr, within 1-2 minutes at 0°C. An appreciable excess of CNBr was required since its decomposition rate in aqueous solutions is extremely fast ($t_{1/2} < 1$ min). Furthermore, restriction endonuclease treatment (*Eco*RII) of the duplex product revealed that the chemically ligated complex was cleaved with the same efficiency as a duplex obtained by enzymatic condensation with T4 DNA ligase, confirming that the oligonucleotides were not modified by the CNBr in any way during the course of the coupling reaction. It was unclear at the time whether CNBr itself activated the phosphate group or if it formed an intermediate with a buffer component as a first step. Mechanistic studies, confirmed that ligation only proceeded efficiently in the presence of tertiary amines such as N-methylmorpholine and 2-(N-morpholino)ethanesulfonic acid (MES). The amine's role was more than just as a pH regulator, and instead, abstracts the cyano group from CNBr, and transfers it to the phosphate group present at the nicked junction thereby effectively activating it for condensation.[271,310]

Knowing all of this, the chemical ligations of all nicked DNA dumbbells were conducted at an optimal concentration of 10^{-4} M phosphorylated nicked precursor (**3.6**-**3.10**) using either N-cyanoimidazole (CNIm)[276] or cyanogen bromide (CNBr)[258,271] as nick-sealing reagents. Low concentrations of oligonucleotides promoted intramolecular complexation and juxtaposed the phosphate-hydroxyl termini in the appropriate spatial orientation for ligation. Samples were dissolved at the required concentration in the appropriate MES buffer for either CNIm or CNBr-induced ligations, denatured by heating at high temperature and cooled slowly to encourage self-association. Reactions were initiated by the addition of either 100 mM aqueous CNIm or 5 M CNBr in dry acetonitrile. The samples were effectively precipitated in quantitative yields from the reaction medium using 2% $LiClO_4$ in acetone. Prior to gel analysis, the ligation products were dissolved in denaturing gel loading buffer, briefly heated at 95°C so as to denature the self-complementary regions of the dumbbell and partitioned on a 20% denaturing gel (8.3 M urea).

The position of the nick and the nature of the reacting nucleotide residues are important for the realization of high nick-sealing efficiency in DNA dumbbell precursors. Moreover, localization of the participating phosphate group at either the 5'- or the 3'-end of the substrate, has been described as a key factor influencing the efficacy of chemical ligation. Nicked dumbbells **3.6** and **3.7** (**Table 3.1**) comprised of identical base sequences, though the terminal phosphate was placed at either the 5'-end (**3.6**) or at the 3'-end (**3.7**). Ligations were conducted using CNIm and left incubating at 4°C for a 2-20 h period. Gel analysis of the resultant products showed the appearance of a single new product band in each case, which migrated faster than the corresponding nicked precursor on a denaturing polyacrylamide gel (**Figure 3.7A** & B). Although the mass/charge ratio

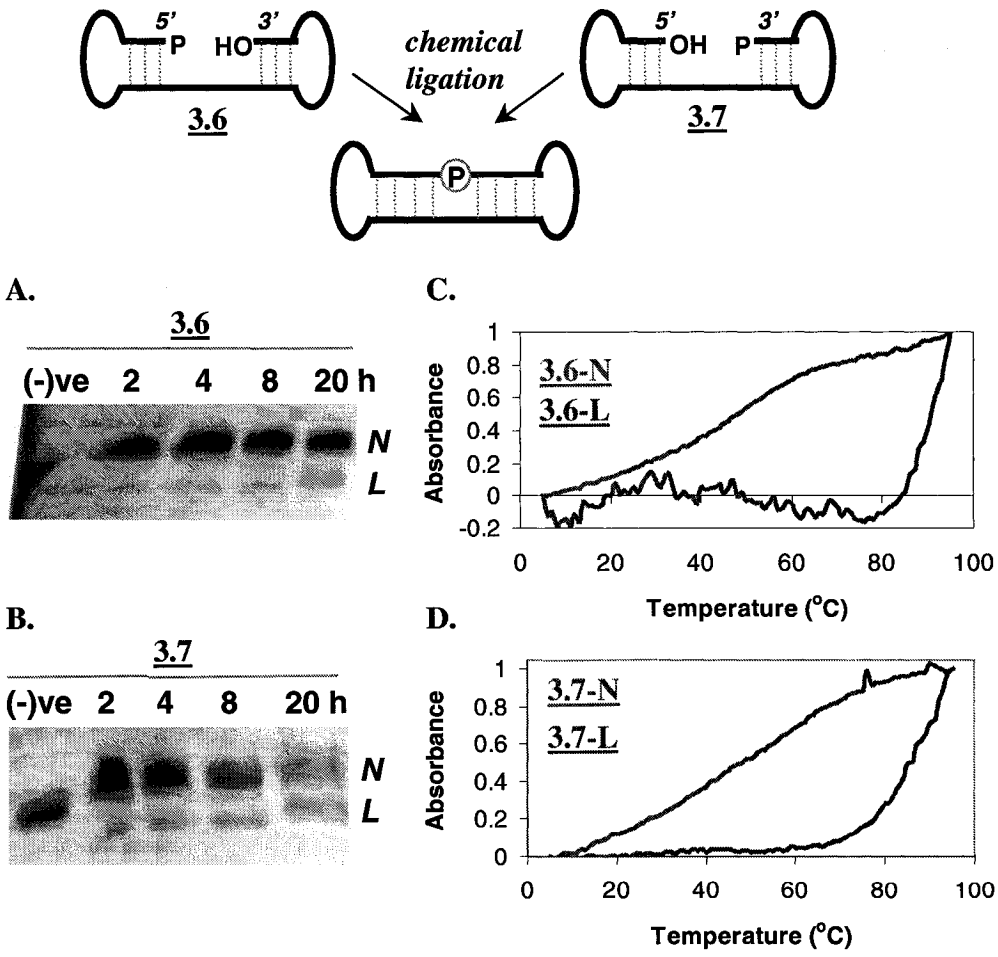


Figure 3.7: Demonstration of the efficiency of 5'-P/3'-OH (**3.6.**) *vs.* 5'-OH/3'-P (**3.7**) ligation using N-cyanoimidazole (CNIm). A & B: 20% PAGE (8.3 M urea). Figure 1B-1D: T_m analysis of nicked (*N*) and ligated (*L*) dumbbells. Buffer: 10 mM Tris-HCl pH 7.5, 10 mM NaCl.

of the nicked to ligated compounds remains nearly indistinguishable, the accelerated migration is conceivably the result of the more globular and compact nature of the circularized products, thereby slipping through the pores of the gel matrix with less resistance.[270,311] Our findings corroborate previous reports that ligation yield improves when the precursor contains a 5'-hydroxyl and a 3'-phosphate, rather than the alternative configuration (**Table 3.2**), however, the yields obtained were significantly lower than desired.[269,303] Longer incubation periods did not appear to improve the yield of circularized product, and instead reaction artifacts were found to accumulate at the top of the gel with increasing time (data not shown). It has been suggested that the major limitation in using 5'-phosphates in such reactions is that once they are activated, their greater conformational flexibility exposes them to hydrolysis by competing water molecules.[269,310] Furthermore, the increased nucleophilicity of the primary 5'-hydroxyl group over the secondary 3'-hydroxyl group is evidently sufficient to overcome any competing reactions and enhance the ligation yield.

In comparing the T_m profiles for the nicked and ligated dumbbells **3.6** and **3.7**, a significant enhancement in the thermal stability ($\geq 30^{\circ}C$) of the ligated substrate is disclosed in both instances (**Figure 3.7 C & D**). Melting of the affixed dumbbell structure generates a covalently closed, single-stranded, circular final state. This increased melting temperature is consistent with previous studies on dumbbell structures, all of which exhibit significantly altered melting.[269,282] The ligation-induced increase in thermal stability is recognized as being exclusively enthalpic in origin.[282] Appropriately, monitoring of the increase in thermal stability proved to be an effective tool for characterizing the ligated structures. All nicked precursors and ligated dumbbells were further characterized by negative MALDI-TOF-MS, demonstrating the loss of a single water molecule upon condensation of the reactive functional groups (Experimental Table 7.2).

Despite the slight increase in circularized product observed with the 3'-phosphorylated substrate (**3.7**), optimization of the product yield was crucial. Besides the localization of

the phosphate and hydroxyl group at the nicked junction, the reaction efficiency also strongly depends on the nature of the nucleotide residues to be ligated. Indeed, the proficiency of the phosphodiester bond formation validates a strong dependence on the local DNA structure at the nick site, and decreases in the following order: 5'-pyrimidine/3'-pyrimidine > 5'-purine/3'-pyrimidine > 5'-pyrimidine/3'-purine > > 5'-purine/3'-purine, with 5'-T/3'-pT bearing the most productive contact at the ligation site.[312,313] When the sequence was modified such that the nucleotide conjointment contained a TpT (**3.8**) rather than a GpT contact (**3.7**), the ligation yield increased by *ca.* 56% (**Table 3.2; Figures 3.7B** and **3.8A**). It is worth mentioning that all the ligations described so far were conducted at a nicked junction directly opposite the 2'-rA insertion. We imagined that this slightly destabilizing insertion may be altering the local base-stacking interactions surrounding the ligation site, and as such, condensation potency should improve if the TpT junction were placed at a site more remote from the 2'-rA substitution. Sequences **3.9** and **3.10** were designed to contain identical base composition so that we could directly compare the effect of ligation directly across from (**3.9**) or distal to (**3.10**) the 2'-rA insert (**Table 3.2**). Furthermore, in order to accommodate this change in sequence design, additional base pairs were introduced into the stems of the dumbbells. The new sequences now had [7+4] and [4+7] complementary nucleotides respectively in the template region (compared to [4+4] nucleotides for **3.8**). Theoretically, this alteration in base-pair abundance should increase the thermal stability of the double hairpin complexes. Evidently, this was indeed the case for **3.9**, whereby the T_m increased by 8°C compared to **3.8**, yet the T_m of **3.10** remained uninflected (54°C). These results allude to the fact that the 2'-rA has a more detrimental destabilizing effect at the center of a duplex region (**3.10**), rather than at the end (**3.9**). Justifiably, this may be the result of end-fraying or "breathing" of nucleotide units which typically results in the destabilization of terminal portions in duplex regions. While this may potentially be the case, it seems that this scenario could be used to our advantage. Since the local geometry of the duplex appeared to be distorted in the region encompassing the 2'-rA unit, then better base-stacking should be achieved further away. Indeed this is what we observed, and the ligation yield improved by as much as 10% for compound **3.10** as compared to **3.9** signifying an improvement in the stacking interaction and base-pair alignment of the TpT

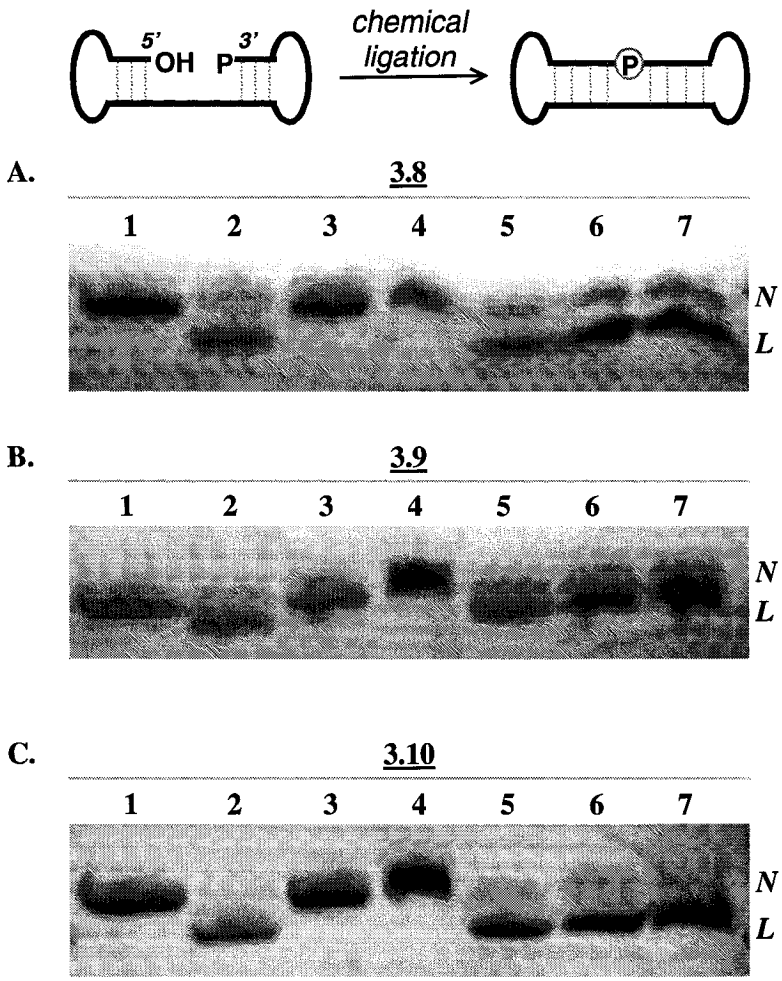


Figure 3.8: 20 % PAGE (8.3 M urea) analysis of the chemical ligation of dumbbells **3.8** (A), **3.9** (B) and **3.10** (C). Lane 1: negative control (nicked precursor alone); Lane 2: CNIm, 4h; Lane 3: CNBr only (no MES buffer); Lane 4: 0.25 M MES only (no CNBr); Lane 5: 0.25 M MES and CNBr; Lane 6: 0.5 M MES and CNBr; Lane 7: 1.0 M MES and CNBr. *N*=nicked precursor; *L*=ligated product.

junction (**Table 3.2**; **Figures 3.8B** and **C**). In all cases, the nature of the condensing reagent, either N-cyanoimidazole or cyanogen bromide, did not appear to be an important factor in optimizing the dumbbell ligation, as reaction extent was similar for both reagents (**Figure 3.8A-C**; Compare lanes 2 & 5-7). Nevertheless, CNBr induced ligation reactions were clearly advantageous as the reaction time was substantially decreased (5 min. *vs.* 4 h). Consistent with results obtained from Shabarova's group and the mechanism of CNBr induced ligations described previously, reactions were contingent upon participation of

the tertiary amine, MES (**Figure 3.8A-C**; Lanes 3 and 5).[271] While they also reported that maximal ligation yields were attained with higher concentrations of MES, enhanced condensation was not apparent in this study when 0.5 M or 1 M MES was utilized (**Figure 3.8A-C**; Lanes 5-7). Extraction of the newly formed product bands from the denaturing gels, and monitoring of their thermal denaturation patterns revealed that the dumbbells were indeed closed, circular entities as demonstrated by the marked increase in T_m (**Figure 3.9**) from the unligated to the ligated species ($\Delta T_m \geq 26\text{-}30°C$). Nicked and closed dumbbells **3.8**-**3.10** were further characterized by MALDI-TOF-MS (**Table 7.2**) and anion-exchange HPLC resolution of reactant and product peaks. A representative chromatogram for dumbbell **3.10** is demonstrated in **Figure 3.10**. Confirmation of the ligated species was conducted by co-injecting a small amount of purified, nicked dumbbell with the crude ligation reaction mixture (**Figure 3.10C**). Under standard conditions (**Table 7.1**), we were capable of effectively separating both the open and closed forms of the dumbbells, even though the phosphate charge difference was only 1 unit. Presumably, the sinuous shape of the ligated species also plays a role, and averts column retention to a greater extent than it's linear counterpart. Using this method, we were thus able to ascertain the extent of condensation by integration of the precursor and resultant product peaks from the obtained chromatograms.

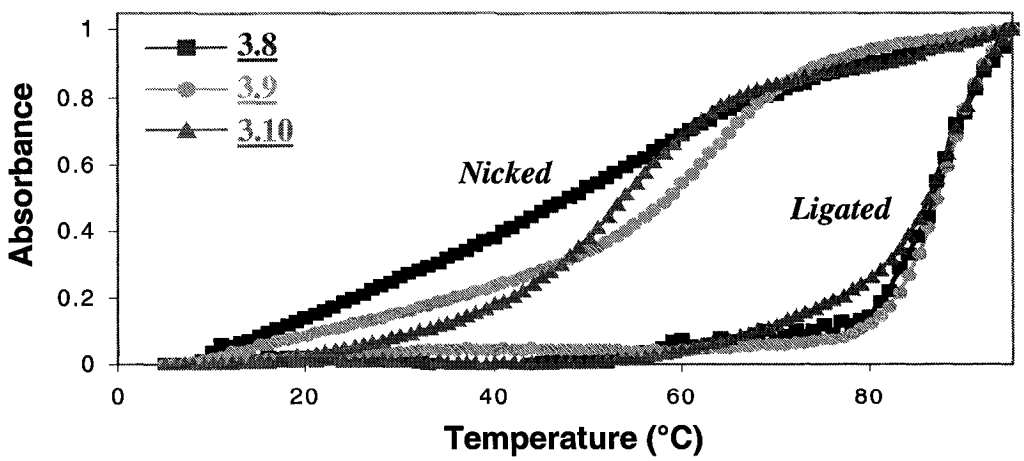


Figure 3.9: T_m analysis of nicked and ligated dumbbells **3.8**-**3.10**. Buffer: 10 mM Tris-HCl pH 7.5, 10 mM NaCl.

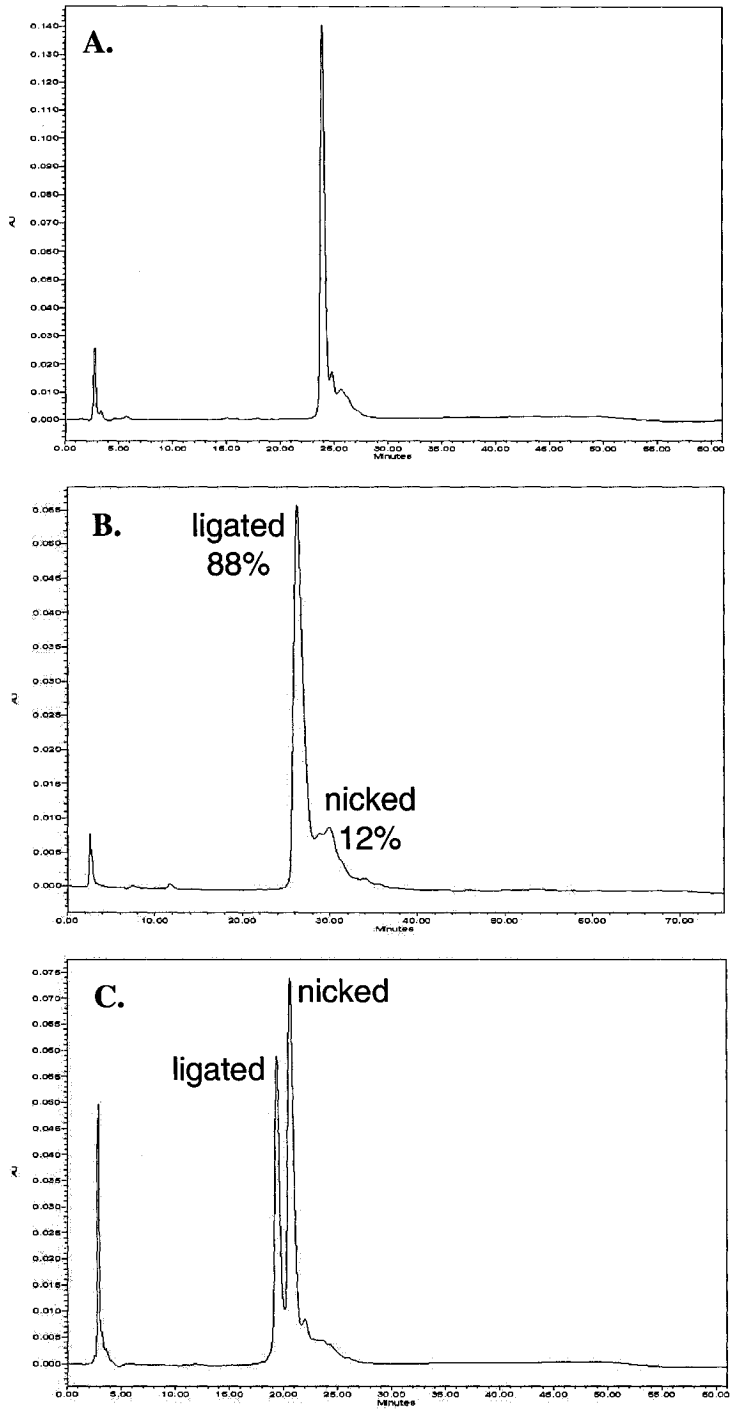


Figure 3.10: Anion-exchange HPLC analysis of nicked and ligated dumbbell **3.10**. Conditions: 10-20% $LiClO_4$ (1M) over 60 min, T=60°C. Panel A: gel-purified, nicked DNA dumbbell **3.10**. Panel B: CNBr ligation reaction mixture of **3.10**. Panel C: CNBr ligation reaction + pure nicked **3.10** (co-injection).

3.2.5. Regiospecific Synthesis of Branched DNA Lariat Dumbbell Precursors

Branched DNA lariat dumbbell precursors **3.11** and **3.12** (**Table 3.1**) were synthesized using previously published methodologies for the regiocontrolled and divergent synthesis of branched DNA with some modifications.[212,221] The key functional moiety in the synthesis is the branching nucleotide 5'-O-dimethoxytrityl-2'-O-*tert*-butyldimethylsilyl-(N^6-benzoyl adenosine)-3'-O-(β-cyanoethyl) N,N'-diisopropyl phosphoramidite or 3'-rA phosphoramidite, which enables the sequential extension of the DNA chain from the 3', 2' and 5' directions, thereby producing Y-shaped oligonucleotides (**Figure 3.11**). The synthetic approach is highly advantageous as it necessitates the use of only commercially available phosphoramidite building blocks and chemistry. The branched (Y-shaped) lariat precursors **3.11** and **3.12**, were constructed to contain regions of self-complementarity in the 5' and 2'-extensions from the rA branchpoint, such that they were capable of intramolecularly folding into a "dumbbell" complex. This in turn would juxtapose the phosphate and hydroxyl termini in the appropriate spatial geometry for effective ligation. The 3'-appendage in both cases consisted of homopolymeric deoxythymidine sequences, and as such is prohibited from forming any type of stable complex with either the 5' or 2'-arms. Such a design would ward off the possible formation of the undesired 3'-linked lariat regioisomeric species. The dumbbell portion of **3.11** consisted of the identical base sequence as **3.7**, while **3.12** had the identical base sequence to **3.10**, which was determined to be our optimal substrate for ligation.

In order to ensure that adequate orthogonal extension was realized, synthesis was initiated on 5'-DMT-thymidine CPG, that contained a controlled, low-loading (*ca.* 5-10 μmol/g) of nucleoside so as to maintain an adequate distance between growing strands.[233] Assembly of the linear 3'→5' DNA strands of **3.11** and **3.12** (**Table 3.1**) utilized standard 3'-DNA phosphoramidites, coupling conditions and reaction times. The 5'-terminus of the linear construct was acetylated with acetic anhydride in order to 'cap' the reactive hydroxyl so as to prevent chain elongation from this point in the molecule during the branch extension step. A small amount of CPG was removed from the column at this point for analysis, and deprotected under standard conditions (*i.e.* linear failure control).

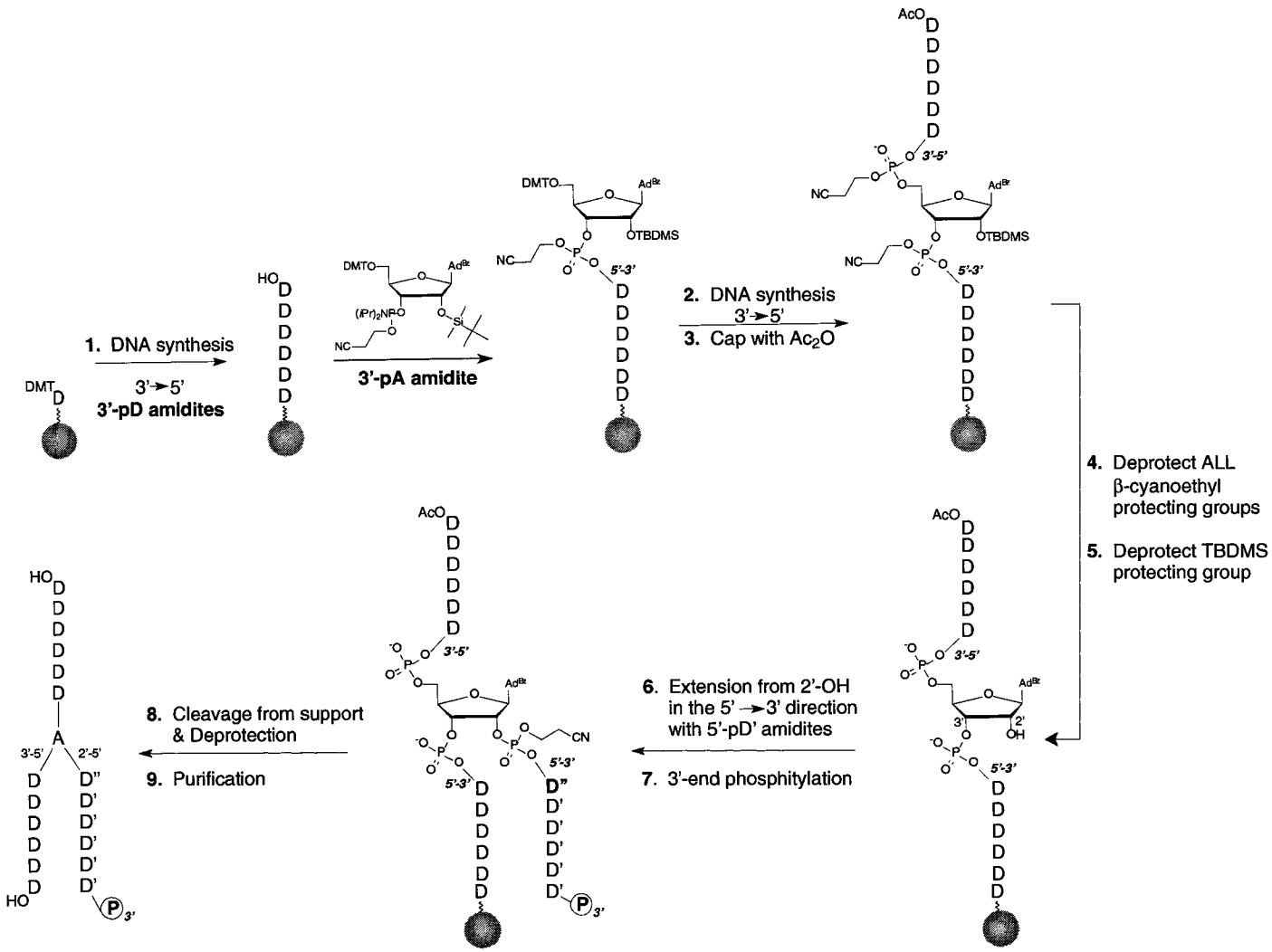


Figure 3.11: Schematic representation of the divergent and regiospecific synthesis of branched DNA. The branching synthon is a standard RNA 3'-phosphoramidite (3'-pA amidite). In the figure above; 3'-pD = standard DNA 3'-phosphoramidites; 5'-pD' = inverted DNA 5'-phosphoramidites; 5'-pD'' = higher concentration (0.3 M) of inverted DNA 5'-phosphoramidite coupled to the 2'-hydroxyl of the ribose branch point; P=terminal phosphate.

Subsequently, all the β-cyanoethyl phosphate protecting groups were eliminated using a manual basic treatment with 2:3 triethylamine/acetonitrile thus converting the highly labile phosphotriester moieties into a more stable phosphodiester backbone.[197,198] Brief treatment (10 min.) with fluoride ions (1 M TBAF in THF) effected the selective removal of the 2'-O-TBDMS protecting group without cleaving the oligomers from the surface or dissolving the support itself. Extended treatment with TBAF results in the cleavage of the succinyl linkage and concomitant eradication of the support-bound oligonucleotide.[212] A thorough rinsing step with THF followed by dry acetonitrile ensured that any excess fluoride ions were washed away. The CPG was then dried thoroughly and reinstalled on the synthesizer. Chain extension from the 2'-hydroxyl of the branching nucleotide (rA) dictated the use of inverted 5'-DNA phosphoramidites. Given that the secondary 2'-hydroxyl of the branching residue is a poorer nucleophile than a primary 5'-hydroxyl, and is located in a sterically hindered position, a higher concentration (0.3 M) and longer coupling time (30 min.) of the first phosphoramidite monomer addition [5'-p(dC) in both cases] was required for adequate coupling.[219-221] Under these conditions, the coupling was still exceedingly low, only 50-60%, suggesting that the 2'-TBDMS group had not been effectively removed. The remaining 2'-extension was elongated with 5'-DNA phosphoramidites at a 0.15 M concentration, a concentration typically used for the incorporation of RNA phosphoramidites. Ultimately, the 3'-terminus of the 2'-extension was phosphitylated with the chemical phosphorylation reagent (see section 3.2.2). The reagent was used at a 0.1 M concentration and coupled under standard DNA phosphoramidite reaction conditions (t=90 s). A small amount of CPG was removed from the column prior to 3'-phosphitylation and deprotected for subsequent analysis (*i.e.* Y-DNA without 3'-phosphate control). The oligonucleotide protecting groups were removed using customary deprotection conditions (see Experimental section 7.4.), and the branched DNAs analyzed by anion-exchange HPLC (Experimental section 7.5.3) and PAGE (Experiemental section 7.5.2).

Analysis of the lariat dumbbell precursor **3.12** by both anion-exchange HPLC and 16% denaturing PAGE revealed the presence of two predominant reaction products; the linear DNA sequence lacking the 2'-branch extension (linear failure), and the fully branched Y-

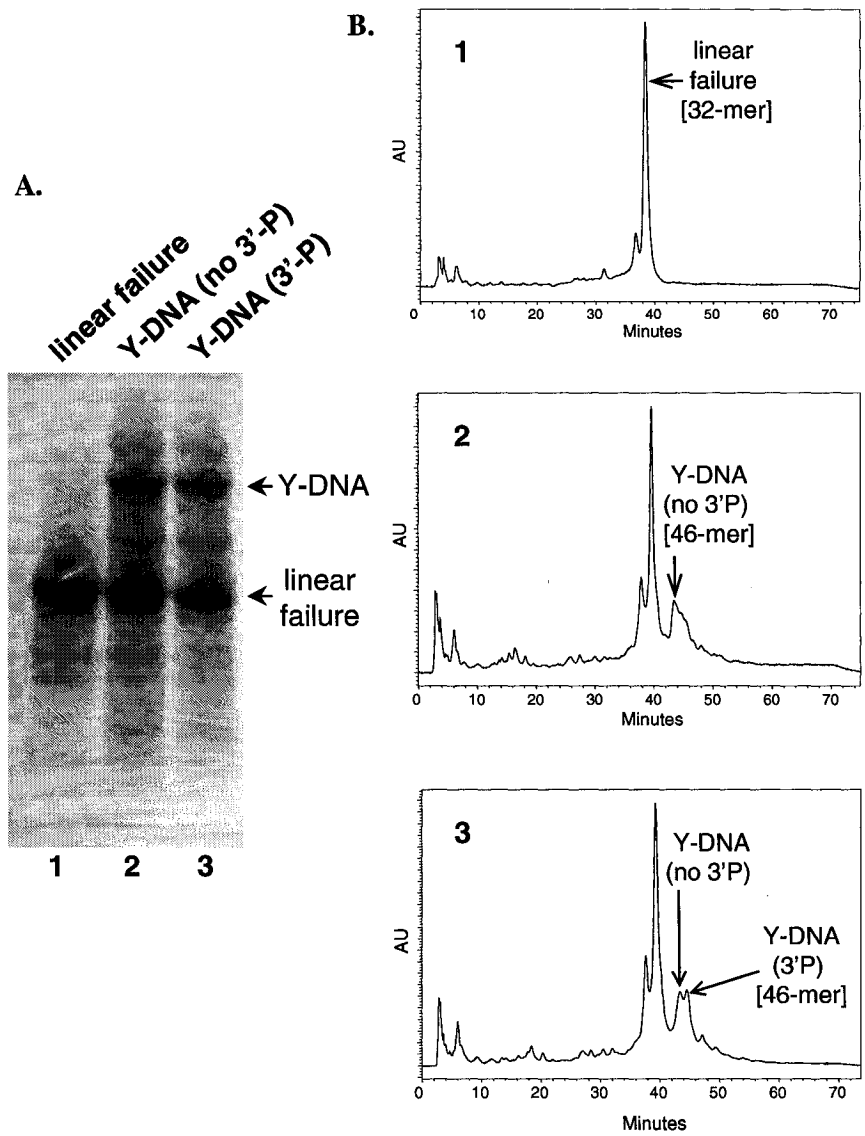


Figure 3.12: Panel A: Analytical 16% PAGE (8.3 M urea) stained with Stains-All® demonstrating the crude synthesis of branched DNA lariat precursor **3.12**. Lane 1: linear failure control DNA (32-mer); Lane 2: crude branched DNA (46-mer) lacking a 3'-phosphate; Lane 3: crude branched DNA (46-mer) **3.12** containing the desired 3'-phosphate. Panel B: Anion-exchange HPLC profiles of the linear failure (1), branched DNA lacking a 3'-phosphate (2) and branched DNA with a 3'-phosphate (3; **3.12**).

DNA structure (**Figure 3.12**). Chromatography of the linear failure control indicated that coupling steps during the synthesis of the linear strand consummated with great efficiency as demonstrated by the absence of significant failure sequence peaks trailing behind the predominant peak (**Figure 3.12 B-1**). Evidently, the low coupling efficiency mentioned

previously, and the significant amount of residual linear DNA present on the gel, clearly signify that chain extension from the branching adenosine subunit was not as successful (**Figure 3.12A-3**). The amount of branched DNA was approximately 35-40% of the total DNA present in the mixture. Whereas the non-phosphorylated Y-DNA chromatogram demonstrates only one new product peak (**Figure 3.12B-2**), the phosphorylated Y-DNA chromatogram indicates the presence of two closely resolved peaks (**Figure 3.12B-3**), one of which is the unphosphorylated species. This suggests that the final phosphitylation step during the Y-DNA synthesis produced incompletely phosphorylated 3'-termini (*ca.* 50%). Likely, this is the result of the increased sterics of extending the 2'-chain in the vicinity of the solid-support during Y-DNA synthesis. The band corresponding to the Y-DNA phosphorylated oligomer (**Figure 3.12A**; Lane 3) was excised from a preparatory 16% denaturing gel (8.3 M urea), soaked in water overnight and desalted by SEC. Although the 3'-phosphorylated and non-phosphorylated species were slightly separable under chromatographic conditions, analysis of both products by PAGE (**Figure 3.12A**; Lanes 2 & 3) reveals that they both exhibit indistinguishable migratory rates. As such, the gel-purified branched DNA lariat dumbbell precursor likely consisted of a mixture of both the phosphorylated and unphosphorylated species.

3.2.6. Chemical Ligation and Characterization of Lariat DNA Dumbbells

The purified Y-DNA dumbbell precursors **3.11** and **3.12**, were dissolved in 0.25 M MES buffer at a concentration of 10^{-4} M of branched molecule. The mixture was heated at high temperature (95°C) to dissociate any self-complementary regions, and cooled slowly to 4°C to promote dumbbell association. We anticipated that such a dumbbell complex would have significant potential to form under these conditions, as previous experiments on branched molecules have clearly designated a preferential base-stacking interaction between the branch point adenine and its adjacent 2',5'-linked substituent.[314,315] Ligation reactions were initiated by the addition of 1/10 volume of 5 M CNBr in acetonitrile, placed on ice for 5 minutes and the DNA oligonucleotides subsequently precipitated from the reaction milieu with 2% $LiClO_4$ in acetone. Prior to PAGE analysis,

the ligated samples were dissolved in gel loading buffer, heated to denature any hybridized segments and suitably partitioned on 12% denaturing gels (8.3 M urea). Attempted ligation of **3.11** failed to produce even the minutest amount of condensed product under CNBr-mediated conditions (**Figure 3.13**). Comparatively, the reaction was conducted under CNIm-mediated ligation conditions, however no detectable new product

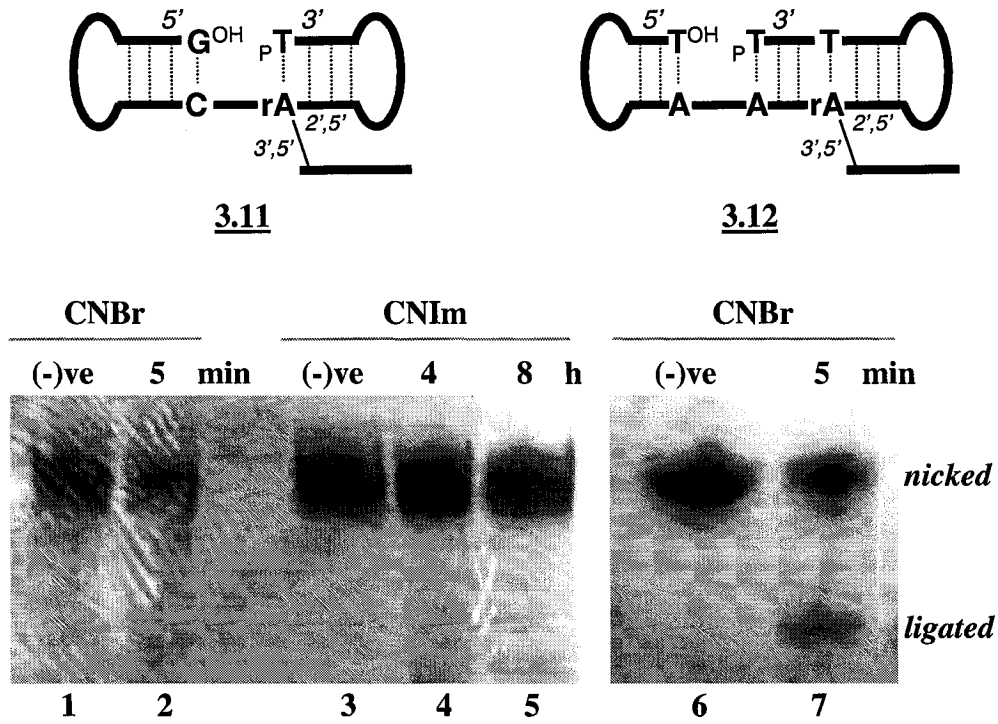


Figure 3.13: 12 % Denaturing PAGE (8.3 M urea) of cyanogen bromide (CNBr) and N-cyanoimidazole (CNIm) mediated ligation reactions on branched DNA dumbbells **3.11** and **3.12**. Lanes 1 & 3: negative control **3.11**; Lane 2: CNBr ligation of **3.11** (5 min.); Lane 4: CNIm ligation of **3.11** (4 hours); Lane 5: CNIm ligation of **3.11** (8 hours); Lane 6: negative control **3.12**; Lane 7: CNBr ligation of **3.12** (5 min.).

was observed, even after 8 h incubation. This was not surprising since we had already demonstrated that the identical dumbbell portion, **3.7**, produced very negligible amounts of ligated material (**Figure 3.7C**). Presumably, this was the result of ineffective contact at the ligation site between the juxtaposed 5'-G/3'-pT as discussed previously (Section 3.1.5). Chemical condensation of **3.12** on the other hand, produced a distinct new product band that migrated faster on the gel than its nicked Y-DNA precursor (**Figure 3.13**). The amount of circularized lariat was lower than anticipated, in comparison to the highly

efficient condensation we had seen for the ligation of its corresponding circular dumbbell **3.10** (**Table 3.2**). One possible explanation for the decrease in ligation yield is that the nicked dumbbell complex was destabilized in the branched configuration compared to its nicked circular counterpart, and as such the ligation junction was not placed in the appropriate spatial arrangement for effective ligation. Alternatively, the co-purification of the unphosphorylated and phosphorylated branched precursors produces a mixture of 5'-OH/3'-phosphate and 5'-OH/3'-OH species at the nick-juncture, and as such, the latter could not be activated and cyclized to form the desired lariat. As such, it is estimated that approximately 75% of the phosphorylated material present in the mixture underwent cyclization.

Clarification of the low cyclization yield was conducted by performing thermal denaturation analysis of the nicked and ligated products of **3.12**. T_m measurement of the hybridized nicked lariat precursor **3.12** exhibited a monophasic transition (**Figure 3.14**) and displayed similar thermal stability to its linear nicked complex **3.10** (ΔT_m=2°C). This

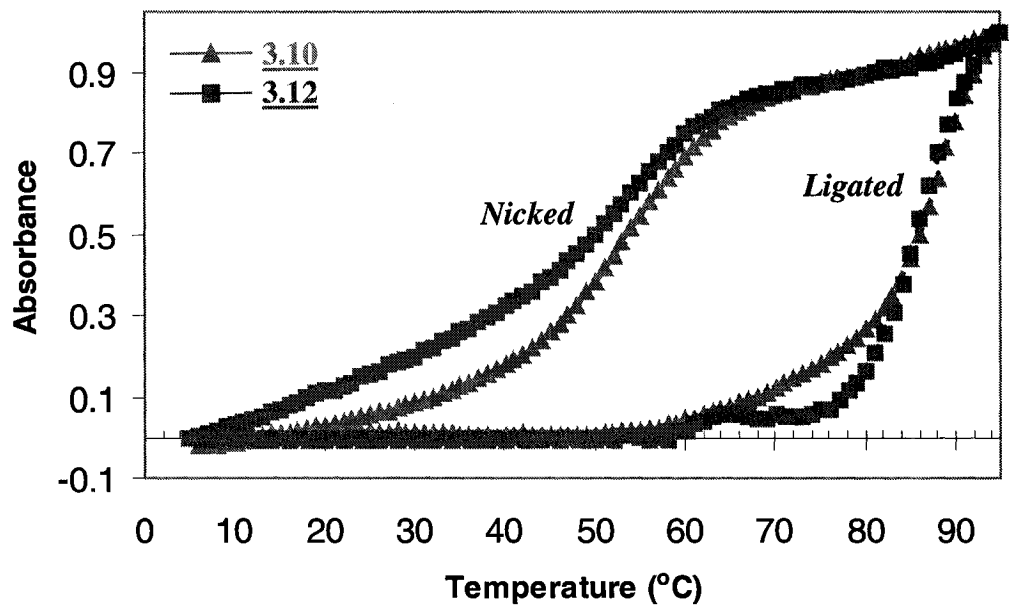


Figure 3.14: Characterization of the ligated dumbbell lariat structure **3.12** by T_m analysis and comparison to its corresponding dumbbell structure lacking the 3'-extension, **3.10**. Buffer: 10 mM Tris-HCl pH 7.5, 10 mM NaCl.

similarity in denaturation profiles indicates that the 3'-extension of the molecule (3'-dT_{16}) does not considerably alter the helix geometry of the dumbbell complex, and instead, likely protrudes out into the solvent, in accordance with a preferential 2',5'-stacking interaction at the branchpoint.[314,315] Confirmation of the desired lariat was accomplished by negative MALDI-MS (see Experimental Table 7.2) and a discernible enhancement in thermal stability of the ligated complex compared with its nicked counterpart (ΔT_m= 34°C; **Figure 3.14**). The resultant T_m data excludes the likelihood that a distorted local helical arrangement was the foundation for the moderate cyclization yields obtained, and instead was solely dependent on the actuality that the lariat precursor contained a mixture of both the inactive unphosphorylated and active phosphorylated species together. To overcome this challenge, employing a higher concentration of the phosphorylation reagent at the 3'-terminus of the branched molecules would insure more complete phosphorylation of the lariat precursor, thus enhancing the cyclization yields. Otherwise, complete phosphorylation would be insured with the use of a functionalized solid matrix (CPG) that releases a 3'-phosphorylated oligonucleotide at the end of a synthesis, yet can withstand the intermediary deprotection conditions required for the regiospecific Y-DNA synthesis.[316,317]

Further characterization of the lariat-dumbbell structures was conducted by incubating the cyclization product with a 5'-exonuclease, bovine spleen phosphodiesterase (BSPDE). Accordingly, both the nicked and ligated forms of **3.12** were radiolabeled at their free 3'-hydroxy termini with [α-^{32}P]-dideoxyadenosine triphosphate (α-[^{32}P]-ddATP) using the enzyme terminal deoxynucleotidyl transferase (TdT) according to product specifications.[318] The labeled oligonucleotides were then treated with spleen phosphodiesterase under its optimal conditions (pH 6.5, 37°C) for 0.5 and 2 h periods. The nicked form of **3.12**, which carries a free 5'-OH, and is therefore a substrate for the enzyme, was almost completely hydrolyzed after the 2 hours incubation period (**Figure 3.15;** Lanes 1-4). The absence of a degradation ladder and the accumulation of one main product band suggests that the enzyme reaches a interruption site at the branchpoint, producing a V-shaped DNA molecule, and is incapable of processing the oligonucleotide

further under these conditions.[187,206] Conversely, the ligated form of **3.12** completely withstood the 5'-terminal processivity of the enzyme, indicating that it was not a substrate for the enzyme as it lacked the presence of a free 5'-terminus (**Figure 3.15**; Lanes 5-8). This confirms that the 5'-end of **3.12** is indeed sealed off in the circular form of a lariat structure. A slight, rapidly migrating degradation band was apparent on the bottom of the gel (not shown). This fast moving band is consistent with 3'-end processivity, a small amount of which contaminates the purified form of the enzyme purchased from the manufacturer (5'-nucleotidase activity < 0.2%; Sigma-Aldrich).

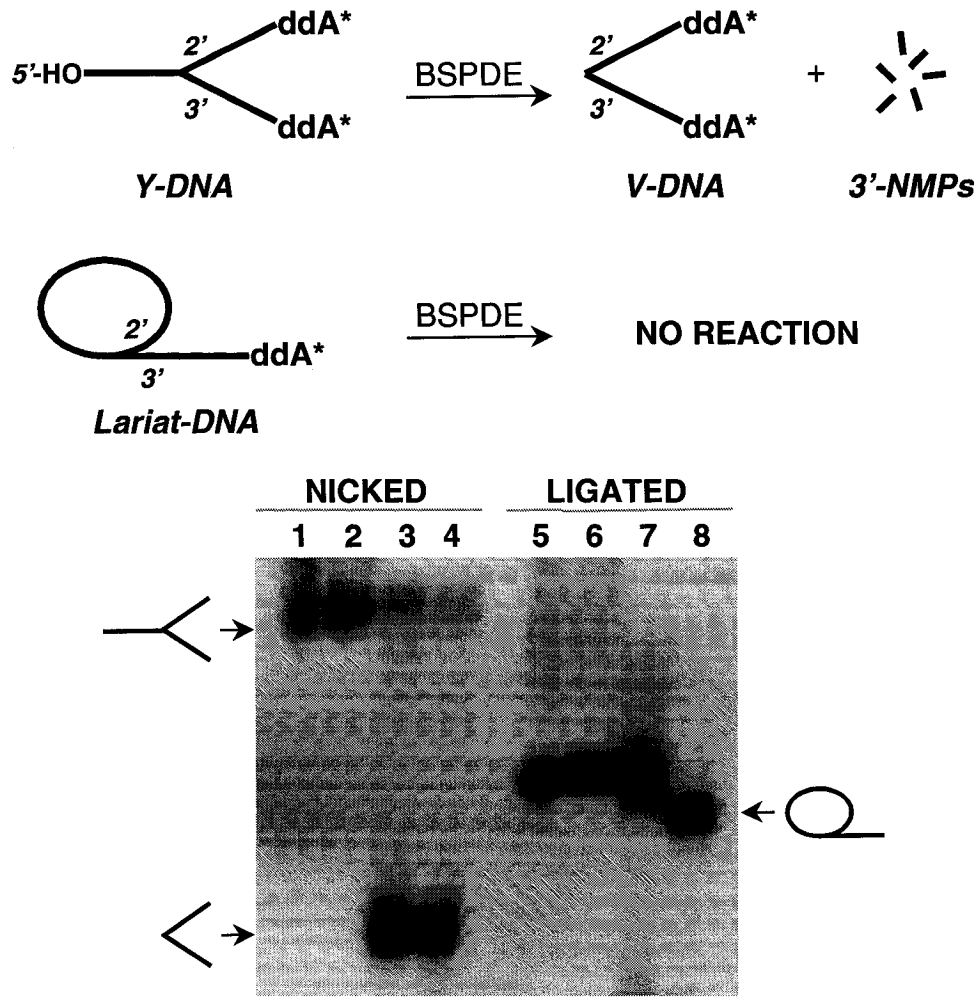


Figure 3.15: Characterization of the nicked and ligated forms of **3.12** using the 5'-exonuclease, bovine spleen phosphodiesterase (BSPDE). 12% denaturing (8.3 M urea) gel. Lanes 1 & 5: negative controls (no buffer + no BSPDE); Lanes 2 & 6: Incubation with buffer only (0.1 M NaOAc, pH 6.5) at 37°C for 2 h; Lanes 3 & 7: BSPDE + buffer (0.5 h); Lanes 4 & 8: BSPDE + buffer (2 h). In the figure above, ddA*=3'-terminal dideoxyadenosine nucleotide radiolabel. NMP=nucleoside 3'-monophosphates.

3.2.7. Global Conformational Analysis of Circular and Lariat-DNA Dumbbells with a 2'-Riboadenosine Linkage

Circular dichroism (CD) is a very powerful, qualitative technique used to probe the global secondary structure of nucleic acid complexes. CD measures the differential absorption of left and right circularly polarized light as a molecule of intrinsic asymmetry is irradiated. Furthermore, it is highly dependent upon both base sequence and base-base stacking interactions within a helix, and as such, the technique is extremely sensitive to conformational changes within nucleic acid secondary structures.[16] Owing to this fact, it can exquisitely differentiate among the diverse helical nucleic acid polymorphs that exist, such as A-form and B-form DNA and any minor distortions within them. It does not, however, provide detailed structural information about the overall conformational attributes of a nucleic acid complex.

Comparison of the CD spectral properties of the linear DNA duplexes **3.1**:**3.3** and **3.2**:**3.3** (**Table 3.1**), demonstrated that the inclusion of a slightly destabilizing 2'-rA insert does not alter the overall duplex structure (**Figure 3.16 A**). The spectral signatures in both cases are indicative of B-form helices, as evidenced by the presence of a positive CD band centered at 280 nm, a negative peak at 255 nm with a crossover between the positive and negative bands at 265 nm, the wavelength maximum for the normal absorption.[15] The positive CD peak and crossover for **3.2**:**3.3** were slightly red shifted with respect to **3.1**:**3.3**, with a small increase in the amplitude of the of the negative CD band. The position and amplitude of the major CD band centered around 280 nm has been qualitatively assigned to the transitions associated with a combination of base stacking and the conformation of the DNA sugar-phosphate backbone.[280] Alternatively, the crossover wavelength has been associated with the thermodynamic stability of base-pair stacking in DNA.[319] This suggests that the slight destabilization by the 2'-rA unit may impose a very small perturbation on the overall helical geometry. The CD profiles are in agreement with comparative CD data obtained for a variety of 2',5'- and 3',5'-linked dinucleoside monophosphates, specifically ApU and ApC.[314] In the case of ApC, the

same linkage found in the substrates used in this study, perturbations between $A^{2'p}C$ and $A_{3'p}C$ were

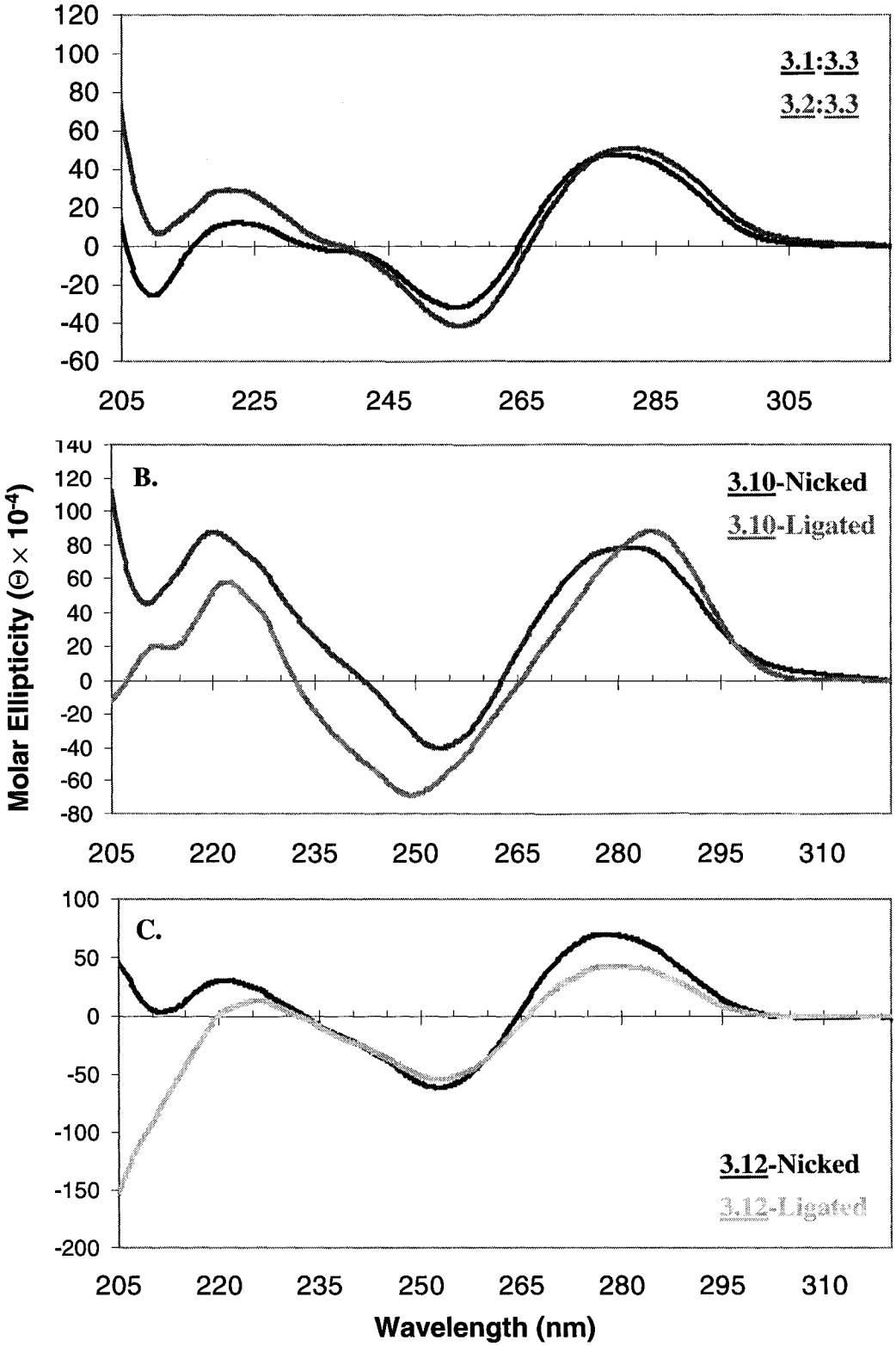


Figure 3.16: Circular dichroic signatures of bimolecular and dumbbell complexes in 10 mM Tris-HCl pH 7.5, 10 mM NaCl buffer. Panel A: Linear duplexes **3.1** and **3.2**; Panel B: Comparison of the nicked and ligated forms of **3.10**; Panel C: Comparison of the nicked branched and ligated lariat forms of **3.12**.

seen in the position of the positive CD and crossover peaks. The difference between the two signatures was attributed to conformational differences between the two compounds.

Regarding the CD signatures of the dumbbell molecules **3.10** and **3.12**, it is evident that in both the nicked and ligated forms, the overall B-type helical geometry is retained (**Figure 3.16 B-C**). The CD contour of the ligated form of **3.10** demonstrates that the positive CD maxima is red-shifted to 285 nm with respect to its nicked counterpart with a slight alteration in the crossover wavelength. However, the ligated lariat-dumbbell **3.12** retains the identical positive, negative and crossover peaks as its nicked complement, with the only difference being a slight decrease in the amplitude of the positive CD band. Given that positive peak, trough and crossover positions in the CD profiles of the nicked **3.10** and **3.12** were almost completely superimposable, suggests that the relative spatial orientation and immediacy of the 2'-riboadenosine and the base at its 2'-position (dC) was similar, and is consistent with a 2'-stacking interaction occurring at this site.[314] Overall, all the CD signatures demonstrated that dumbbell association and subsequent ligation produced B-form DNA structures with negligible amounts of distortion in the helix imposed by the 2'-rA linkage.

3.3. Intermolecular Template-Directed Chemical Ligation of DNA and RNA Lariats

3.3.1. Experimental Design and Project Objective

Although the previous design for lariat synthesis based on the dumbbell strategy affords regioisomerically pure lariat DNA molecules with reasonably high yields, it is limited by the fact that an intramolecular template is a requisite trait, and as such, lariat synthesis is restricted to those sequences that contain self-complementarity within the loop structure itself. In order to potentially synthesize lariat molecules of any base sequence as well as more biologically relevant RNA lariat molecules, we looked towards the complementary template-mediated chemical ligation of branched oligonucleotides using an

intermolecular template guide sequence or "splint". Complementary splint oligonucleotides (DNA or RNA) were constructed such that they were capable of forming a stable complex (*i.e* intermolecular duplex) with the terminal regions of the 2' and 5' extensions of a branched nucleic acid, either DNA or RNA. Previous data concerning chemical ligations allowed us to predict some favorable properties with respect to designing appropriate sequences in the complementary region of the bNA.[258,271,303,310,312] Primarily, the lengths of the nicked oligonucleotides should be sufficient to impart stable duplex formation with its complementary template oligonucleotide. Effectively, this would align the reactive phosphate and hydroxyl junction for covalent ligation with the condensing reagent cyanogen bromide (CNBr), thus forming a circular oligonucleotide with a 3'-tail (**Figure 3.17**). Ideally, the precursor would contain at least 6 + 6 nucleotides complementary to the DNA so as to meet this requirement.[258] Furthermore, the loop size has to exceed the number of nucleotides involved in base-pairing with the template (> 12-nt), in order to be sure that the ligation site is completely hybridized to the splint, and that the template does not form a complex all the way around the loop, as such a duplex might be too distorted.[311,320] Loop sizes that are not at least twice as long as the template region have been shown to result in the construction of concatameric oligomers rather than circular ones.[258] Explicitly, the nicked precursor would preferentially bear a 3'-phosphate group and free 5'-hydroxyl as described in section 3.2.4, as well as the most productive contact at the internucleotide junction, namely 5'-T/3'-pT. In addition, both the template region and complementary splints were designed to bear mirror sequence symmetry, thereby minimizing the number of aggregate products that could potentially form.

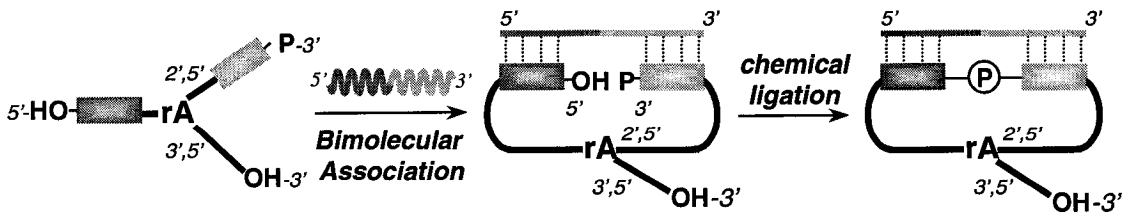


Figure 3.17: Bimolecular association of regiospecific branched DNA with a complementary DNA splint for the synthesis of lariat DNA.

Branched lariat precursors contained a riboadenosine unit at the branchpoint, in keeping with the natural lariat structure. Terminally phosphorylated branched DNA (Y-DNA) lariat precursors were synthesized using the regiospecific and divergent synthetic methodology described previously for the synthesis of lariat-DNA dumbbell precursors, however, they did not include any form of extensive sequence complementarity within the molecule (**Figure 3.11**). Alternatively, branched RNA molecules were synthesized using the convergent solid-phase methodology and the adenosine 2',3'-O-bis-phosphoramidite branching monomer **1.1** as described in Chapter 2. Phosphate groups were introduced at the respective 3'-termini using the chemical phosphorylation reagent as described in section 3.2.2 (**Figure 3.6**).

3.3.2. Intermolecular Template-Mediated Synthesis and Characterization of Lariat DNA

Linear templates (**3.17**) and complementary DNA (**3.13**-**3.15**) as well as the branched DNA lariat precursor (**3.16**) were synthesized on an Applied Biosystems 381A synthesizer using standard β-cyanoethyl phosphoramidite chemistry and commercially available reagents (**Table 3.3**). Oligonucleotides were purified by denaturing PAGE (12-20%, 7M urea), desalted by SEC and their nucleotide composition confirmed by MALDI-TOF-MS and gel mobility comparison to purified DNA standards. As mentioned, branched DNA was erected to contain terminal regions at the 2'- and 5'-extensions of the molecule that could associate with a DNA splint (**3.17**) by simple Watson-Crick complementarity. The complementary termini comprised of 7 nucleotides stretches that could be recognized by the 14 nucleotide template splint on each arm. Adjacent to the terminal portions were two undecanucleotide non-hybridizable linkers appended to the rA branchpoint (**3.16**; **Table 3.3**). Additionally, the 3'-rA extension was a homopolymeric deoxythymidine oligonucleotide that could not hybridize to the splint DNA in any manner, thereby excluding it from the loop structure (**Figure 3.17**). It has previously been shown that the efficiency of circularization of an oligonucleotide correlates well with the secondary structure of the precursor oligomers, with hairpin-like structures possessing free termini favoring the circularization reaction.[258] Preorganization of the

Table 3.3: Sequences of DNA oligonucleotides used in the intermolecular template-mediated synthesis of a DNA lariat.

Code	Sequence (5'→ 3')	Designation
3.13	gcgttgt$_P$	Nicked 5'-DNA Control
3.14	^{HO}tgttgcg	Nicked 3'-DNA Control
3.15	gcgttgttgttgcg	Full Length DNA Control
3.16	^{HO}tgttgcgt$_{11}$A$^{2'5'}$(t$_{11}$gcgttgt$_P$)$_{3'5'}$t$_{20}$	Lariat-DNA Precursor
3.17	cgcaacaacaacgc	DNA Template

Notation: small caps=deoxynucleotide residues; ; A=riboadenosine; OH=terminal hydroxyl; P=terminal phosphate; underlined oligonucleotides represent those which are complementary to the intermolecular DNA template **3.17**.

precursor into a more rigid hairpin-like entity minimizes the entropic penalty of hybridizing two freely rotating complementary sequences to one template in solution.[321] As such, the predicted secondary structure of the loop portion of the lariat precursor was analyzed using the DNA MFOLD server and imposing certain reaction constraints.[287] Computer analysis revealed that the 37-mer loop would likely fold into one probable hairpin-like structure with a negative ΔG at the ligation temperature (4°C), leaving both reactive 5' and 3' termini free to associate with the complementary splint (**Figure 3.18**). The mini-stem portion of the hairpin appeared to be stabilized by two Watson-Crick hydrogen bonds between C_6/G_{33} and G_7/C_{32} while the loop closing base-pairs were stabilized by a non-canonical T_8/G_{31} wobble base-pair. Prior to assessing the efficiency of ligation, the thermal stability of **3.16** complexed with its template splint, **3.17** was gauged. Equimolar amounts of precursor and template were heated in the appropriate annealing buffer, and cooled slowly to allow intermolecular association. Under CNBr ligation conditions (*i.e.* 0.25 M MES buffer) the complex displayed a cooperative, monophasic transition with a respective T_m of 39°C, indicating that the complex was indeed stable enough for ligation (**Table 3.4**). Appropriately, the stability of the complex was also measured in a buffer of lower ionic strength (*i.e.* 10 mM Tris-HCl pH 7.5, 10 mM NaCl). As expected, the T_m of the complex decreases to 12°C under these conditions. Comparatively, the thermal stability of the ternary complex between **3.17** and

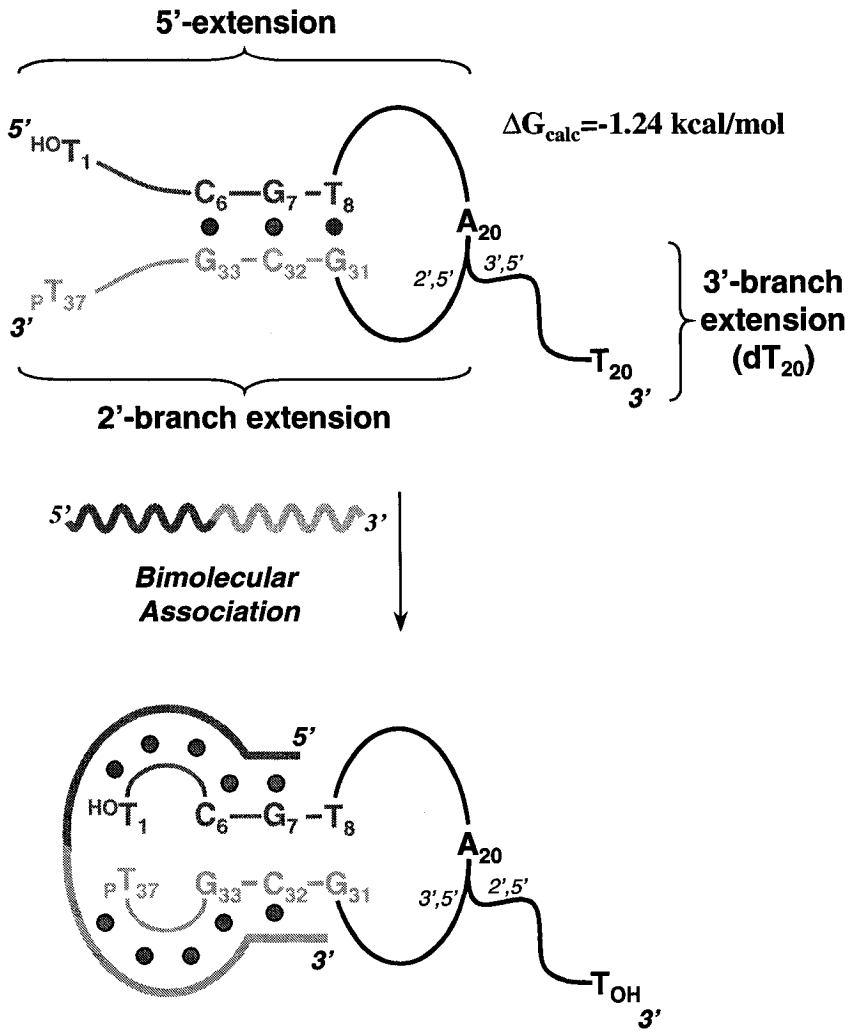


Figure 3.18: The predicted favorable secondary structure of the lariat DNA precursor **3.16** under MES buffer ligation conditions (4°C). Predicted structures were obtained from the DNA MFOLD server available at: http://www.bioinfo.rpi.edu/applications/mfold/old/dna/form1.cgi. The 3',5'-branchpoint T_{20} extension was omitted for calculation purposes. The calculated ΔG is within ±0.06 kcal/mol. The bases in the loop are numbered with respect to their position from the 5'-terminus of the molecule. Blue dots represent standard Watson-Crick (WC) base-pairing while the red dot represents a G-T wobble base pair.

the nicked 5'- and 3'- DNA controls **3.13** and **3.14**, corresponding to the complementary regions of the Y-DNA precursor, was 7°C higher than that of **3.16** (**Table 3.4**; **Figure 3.20**). Dissociation of the ternary complex displayed a single, cooperative transition, given that the hybridized portions were mirror images of one another. This destabilization in the looped complex is consistent with the proposed hairpin-like

Table 3.4: Thermal denaturation and ligation data for nicked and ligated intermolecular DNA complements with their corresponding DNA template splint.

Template	Complement	T_m (°C)		Ligation Yield (%)
		Tris[a]	MES[b]	
3.13	**3.17**	17.0	36.0	
3.14	**3.17**	17.3	33.0	
3.13 + 3.14	**3.17**	19.1	36.5	
3.16 (nicked)	**3.17**	12.0	39.0	
3.15	**3.17**	51.0	-	
3.16 (ligated)	**3.17**	41.7	-	45

T_m values are the average of 3 successive runs and are within ±0.5°C. [a]10 mM Tris-HCl, pH 7.5, 10 mM NaCl; [b]0.25 M MES pH 7.6, 20 mM $MgCl_2$. dash=not determined. The ligation yield was determined by densitometric quantitation of the circularized product band using the UN-SCAN-IT (Silk Scientific) software program.

secondary structure, such that the DNA template has a more difficult time accommodating the complementary strands in the kinked portions of the stem-loop region. Quite possibly, the template is not fully base-paired in this domain, and in effect, end-fraying is a likely consequence. Interestingly, when the nicked controls, **3.13** and **3.14**, were hybridized independently to the DNA splint, their respective melting temperatures were lower than their collectively hybridized value (**Table 3.4**). Previously, Agrawal's group has shown that when short, tandem oligonucleotides bind without a base separation between their binding sites to their target, they bind cooperatively, with a thermal stability greater than the average of the independent duplexes.[322] As such, the cooperative interactions between **3.13** and **3.14** are probably driven by stacking interactions amid the terminal bases, resulting in a highly competent and productive contact.[312,313] Presumably, this analogy holds true for the nicked **3.16** complex, although supporting data regarding the independent associations of the 5'- and 2'-extensions to the template were impossible to obtain given the bimolecularity of the complex.

Chemical ligation of the DNA lariat precursor **3.16** was conducted using a precursor to template (**3.17**) ratio of 1:1 at a total nucleotide concentration of 10^{-4} M. Higher concentrations of splint to precursor (*e.g.* 10^{-3} M) have been shown to lead to the preferential creation of dimeric product species.[258] Subsequent to heat denaturation, the duplex was left to associate slowly at 4°C in order to promote juxtaposition of the terminal 5'-hydroxyl/3'-phosphate junction. Ligations were initiated by the addition of CNBr, and after 5 minutes, the condensation products were quantitatively precipitated from solution. PAGE analysis of the ligation mixture exposed a newly formed, major product band that migrated faster than its nicked precursor, and was formed in approximately 45% yield (**Figure 3.19B**). A minor product band, which migrated near the top of the gel, was also evident (*ca.* 15%). As demonstrated in **Figure 3.19A**, the template splint DNA **3.17** can associate with the branched lariat precursor **3.16** via two different binding modes. At high dilution, bimolecular association prevails to form the desired 57-mer circularized lariat structure consisting of a 37-nt loop and homopolymeric 20-nt stem. The resultant lariat structure likely exists in a more compact form such that it migrates through the cross-linked matrix at a faster rate than its analogous nicked precursor. However, a competing reaction is the termolecular association of the splint DNA with two molecules of precursor, which aligns the junction in a head-to-tail assembly, resulting in the production of a dimeric, hyper-branched species containing 114 nucleotides (**Figure 3.19A**). Normally, such a reaction is prevalent at elevated concentrations of oligonucleotide precursors, suggesting that a higher dilution in our experiments (*e.g.* 10^{-5} M) might favor exclusive lariat formation. The sub optimal ligation efficiency, established by the amount of remaining nicked precursor (*ca.* 40%), suggests that the terminus of the 2'-branch extension is not completely phosphorylated, owing to the divergent introduction of a terminal phosphate group in the vicinity of the solid-support (see section 3.26).

The three representative bands were excised from the gel and their nucleotide composition discriminated by negative MALDI-TOF mass spectrometry (Experimental **Table 7.3**). The molecular weight of the slowest moving band was indeed consistent with dimer formation, exposing a molecular weight of approximately 35

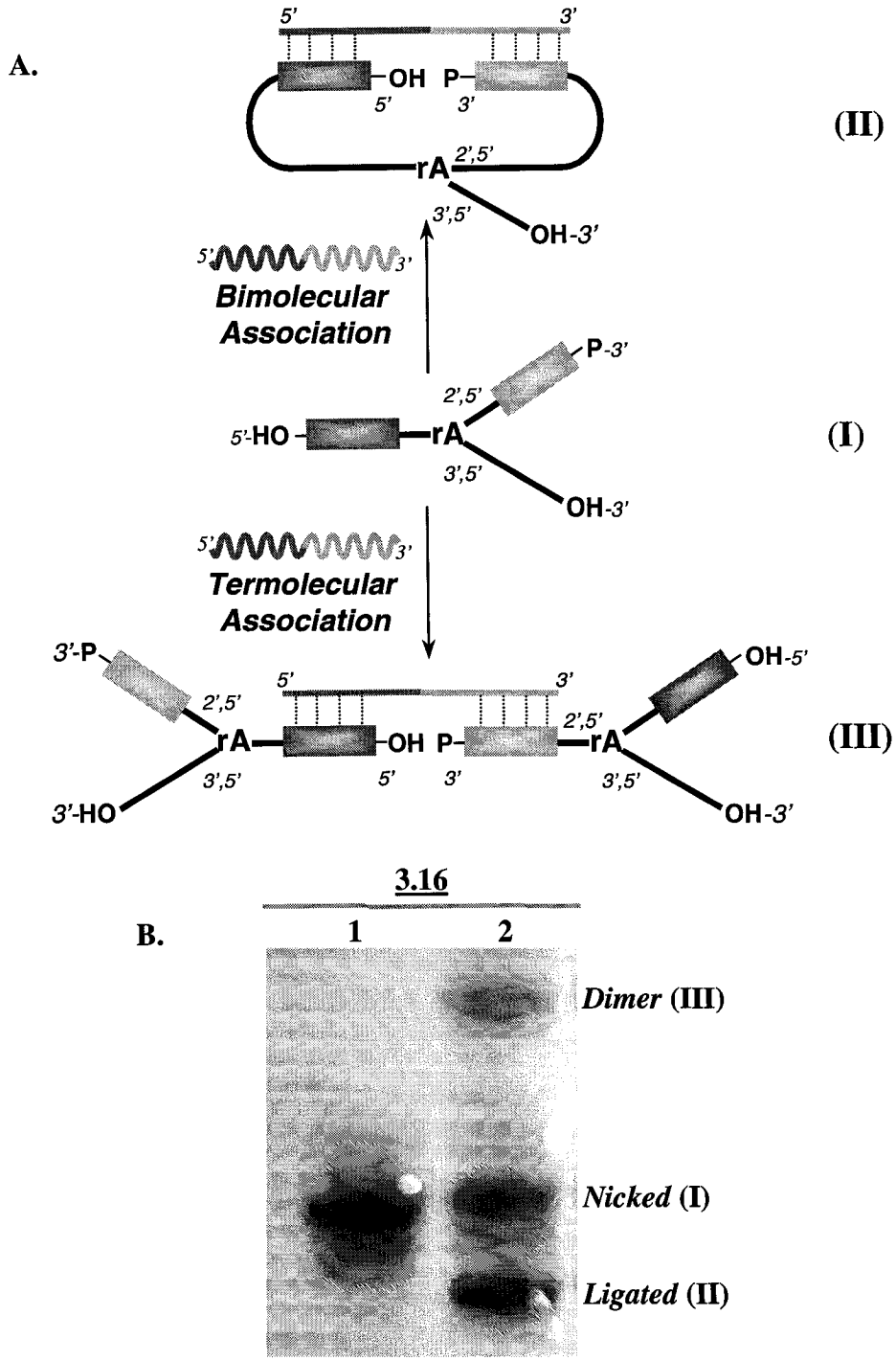


Figure 3.19: Intermolecular association and chemical ligation of branched Y-DNA **3.16**. Panel A: Bimolecular association of the Y-DNA template with complementary splint DNA (**3.17**) produces the desired circularized loop. Competing termolecular association (2 template Y-DNAs + 1 splint DNA) produces dimeric Y-DNA hyperbranched species. Panel B: Analysis of the CNBr-mediated ligation of **3:16** in the presence of 1:1 complementary splint **3.17** using 12% denaturing PAGE (8.3 M urea). Lane 1: negative control **3.16**; Lane 2: CNBr ligation of **3.16** in the presence of template **3.17** (1:1).

KDa. The respective molecular ions of the predominant product band and nicked precursor displayed a difference of 18 units, thereby confirming the condensation of the reactive phosphate/hydroxyl junction and consequent loss of a water molecule. Analysis of the thermal denaturation profiles for the sealed form of **3.16** and its open structure reinforced the covalent circularity of the molecule. In its ligated form, the complementary region of the circle is compelled to form a full 14-nt base-pair with the template **3.17**, thereupon exhibiting an enhanced thermal stability compared to its nicked correlative. Indeed, an enhancement of *ca.* 30°C is disclosed, indicating that a new phosphodiester bond between the nicked junctions had been properly accomplished (**Table 3.4**; **Figure 3.20**). Nonetheless, the T_m of the closed-loop structure complexed to

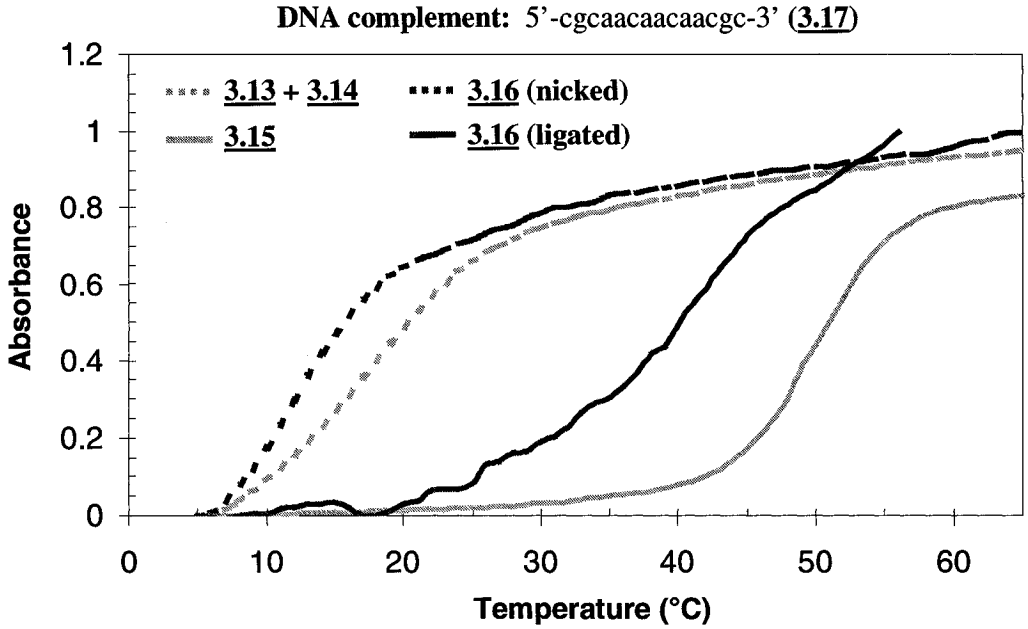


Figure 3.20. Thermal denaturation analysis of nicked and ligated **3.16** and comparison to the nicked DNA complementary region (**3.13** + **3.14**) and full DNA complementary region (**3.15**). All compounds were hydridized to their DNA splint sequence, **3.17** (1:1). T_m's were obtained in 10 mM Tris-HCl, pH 7.5, 10 mM NaCl.

3.17 was lower than that of the relevant full length linear 14-nt complement **3.15** bound to the identical template (ΔT_m=9°C; **Figure 3.20**). Furthermore, the transition appeared to be broader and less cooperative, signifying that full base-pairing between the splint and complementary regions was lacking. The binding of the single stranded target to a

circular complement imposes a topological constraint since duplexes are right handed helices, and each turn of the helix (*i.e.* 10 base-pairs) requires that the strand pass through the circle once.[259,320,321] This requires that the circular portion first be threaded onto the end with the single stranded target and then work its way down to the appropriate binding site. The rigidity of the loop probably causes the 14-nt template to be incompletely hybridized at its terminal regions resulting in significant end breathing and a resultant drop in the overall complex stability.

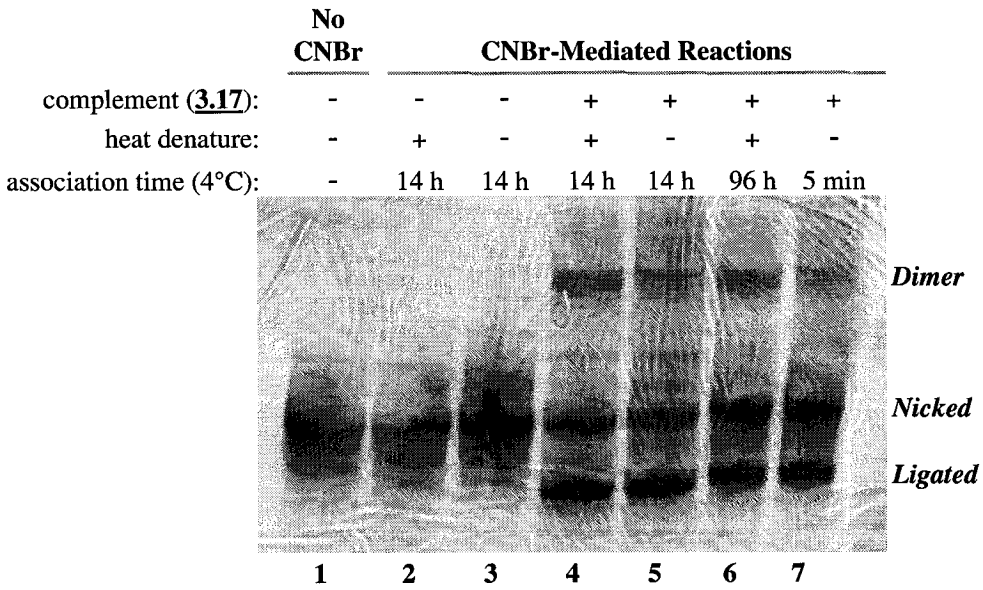


Figure 3.21: Assessment of CNBr-induced ligation of **3.16** under diverse reaction conditions on a 12% denaturing gel (8.3 M urea). Oligonucleotides were dissolved at a total concentration of 10^{-4} M precursor and complement in 0.25 M MES pH 7.6, 20 mM $MgCl_2$ buffer.

The capacity for forming the desired DNA lariat structure under diverse reaction conditions was also assessed by gel electrophoresis (**Figure 3.21**). The results unmistakably established that in the absence of complementary splint DNA **3.17**, none of the circularized product, or dimeric species for that matter, were formed (compare lanes 2 & 4). Moreover, production of the covalently closed circle did not require prior heat denaturation of the precursor-splint mixture before the addition of condensing reagent, suggesting that the precursor was in a freely associative state at the ligation temperature (lanes 4 & 5). Interestingly, the amount of time required for association of the complex

prior to ligation was inconsequential as well. When the complex was left to associate over a 3-day period, the amount of circularized product was nearly identical to that obtained when the precursor and template were merely dissolved in ligation buffer and placed on ice with CNBr for the duration of the ligation time (5 minutes; lanes 6 & 7). Indeed, the complementary region of the loop appears to be in a pre-organized, and accessible conformation for hybridization to its template splint. In addition, the amount of dimerization adduct formed using an extended association time was slightly higher than that produced under kinetic conditions, suggesting that the pre-organized hairpin-like secondary structure loop opens up to allow for termolecular association with the splint. As such, hairpins with free 3' and 5' ends appear to be constructive structures for oligonucleotide circularization, and the extent of terminal ligation most probably reflects the equilibrium concentration of the hairpins in the starting reaction conditions.

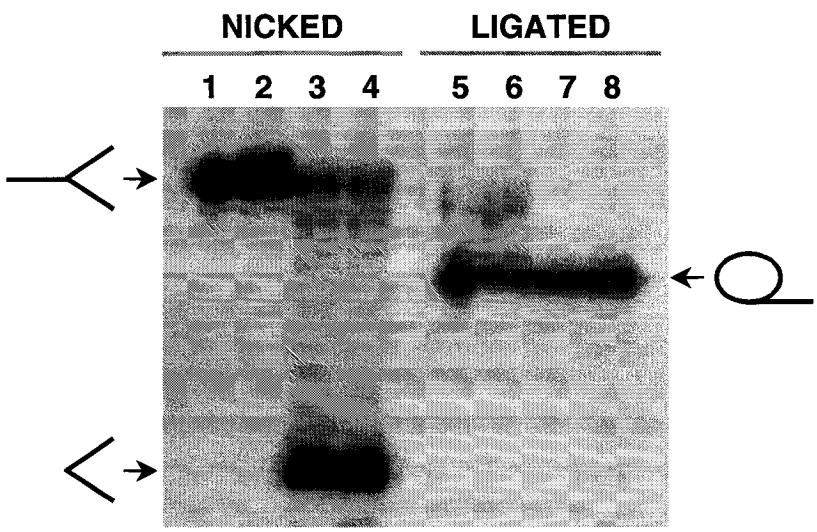


Figure 3.22: Characterization of the nicked and ligated forms of **3.16** using the 3'-exonuclease, bovine spleen phosphodiesterase (BSPDE). 12% denaturing (8.3 M urea) gel. Lanes 1 &5: negative controls (no buffer + no BSPDE); Lanes 2 & 6: Incubation with buffer only (0.1 M NaOAc, pH 6.5) at 37°C for 2 h; Lanes 3 & 7: BSPDE + buffer (0.5 h); Lanes 4 & 8: BSPDE + buffer (2 h).

Further characterization of the closed-nature of the loop was conducted by enzymatic degradation with the 3'-exonuclease, BSPDE (**Figure 3.22**). A radioactive α-[^{32}P]-ddAMP probe was introduced at the 3'-termini of both the nicked and ligated forms of **3.16** by way of the enzyme terminal deoxynucleotidyl transferase (TdT), such that the

degradation pattern could be monitored electrophoretically.[318] As described previously, the unligated loop structure bears a free 5'-hydroxyl terminus, which makes it an apposite substrate for the enzyme's 5'-processivity. Indeed, the nicked form of **3.16** was readily degraded by the enzyme leading to the accumulation of one major product band that corresponded to the branched DNA lacking the 5'-extension (V-shaped DNA; lanes 3 & 4). Contrarily, the ligated form of **3.16** was completely insusceptible to enzymatic cleavage, establishing that after ligation, the 5'-terminus of the molecule is integrated into the sugar-phosphate backbone of the DNA and is no longer accessible to the exonuclease.

3.3.3. Conformational Assessment of Nicked and Ligated Bimolecular DNA Complexes

The global secondary structures adopted by the nicked and ligated forms of **3.16** were investigated by circular dicroism (**Figure 3.23**). CD signatures were obtained at variable wavelengths (200-320 nm) but at constant temperature (4°C) thus ensuring that the oligonucleotides were in their complexed state. Under the low ionic strength conditions used (10 mM NaCl), the single stranded DNA splint **3.17** is not highly structured in solution, however it demonstrates a weak B-type helical signature (**Figure 3.23A**). On the other hand, the spectral profile of the nicked lariat precursor **3.16** indicates that the Y-structure, in its unbound state is B-form as well, but is more structured than the single-stranded splint, given the amplified magnitude of the positive and negative cotton effects. Complexation of the splint **3.17** to its complementary nicked precursor **3.16** results in a CD profile that is strictly B-form in character, and appears to be a simple summation of their individual spectral signatures. The helix type is defined by a characteristic positive cotton effect at 280 nm, a negative peak at 250 nm with a crossover between the positive and negative bands at 265 nm. No perturbation of the helical structure is observed upon binding of the splint to the precursor lariat. (**Figure 3.23A**). When the template is bound to the closed, circular form of **3.16**, a decrease in the amplitudes of both the positive and negative CD signals is disclosed (**Figure 3.23B**). Furthermore, the major CD band centered at 280 nm in the nicked complex is red-shifted by approximately 5 nm in the ligated complex. The decrease in spectral amplitude and the shift in crossover

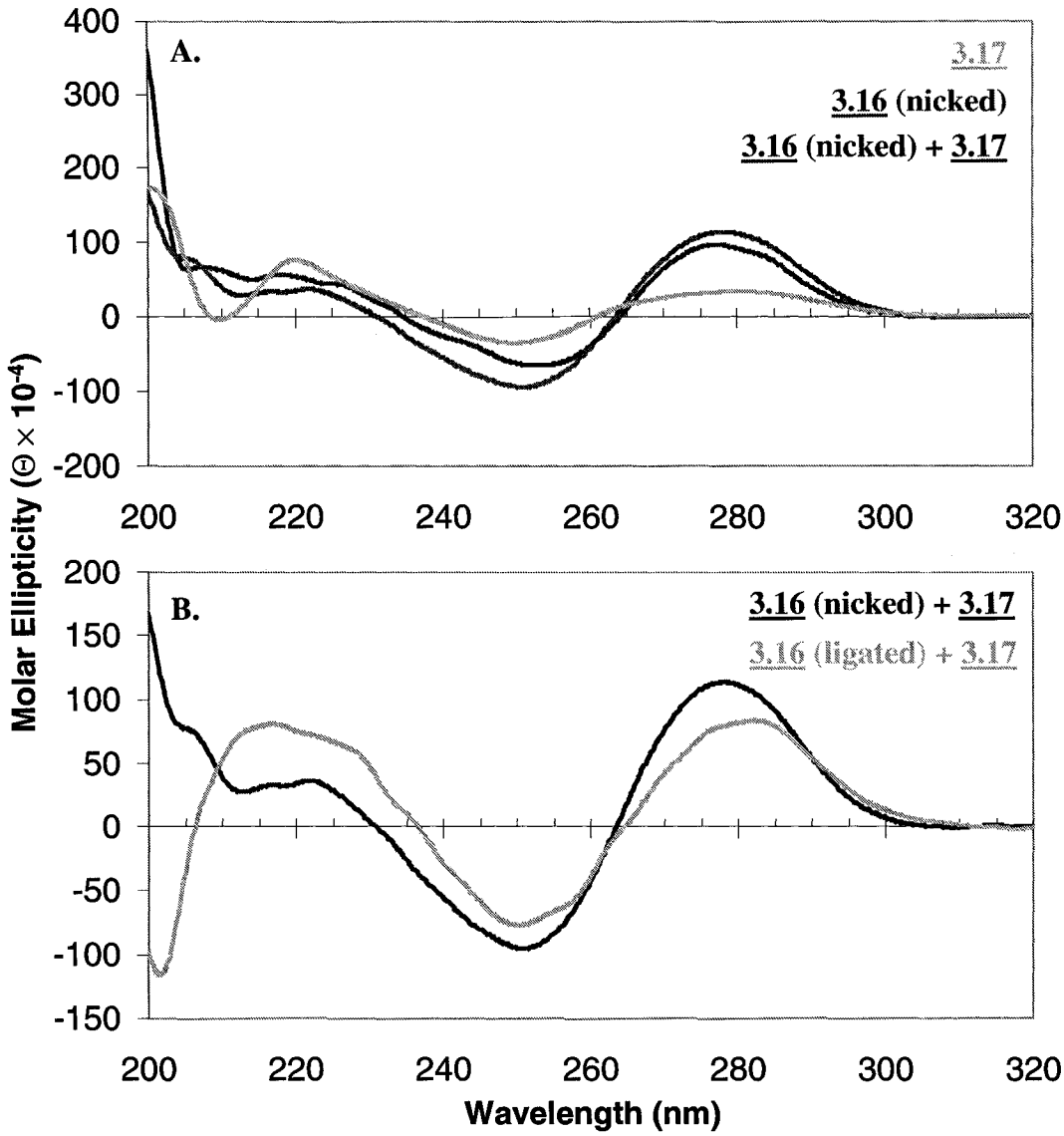


Figure 3.23: Circular dichroic signatures of the intermolecular complexes used in the CNBr-mediated ligation of **3.16**. CD's were obtained in 10 mM Tris-HCl pH 7.5, 10 mM NaCl buffer. Panel A: Comparison of the associated and unassociated nicked forms of **3.16**. Panel B: Comparison of the nicked and ligated forms of **3.16** complexed to **3.17**.

wavelength qualitatively signify that base-stacking in the ligated complex is not as pronounced as in the nicked complex. This suggests that although the template has threaded through the loop and is bound to its complementary region, portions of local disruption, likely at the terminal regions of the complex are found along the helix axis.

3.3.4. Intermolecular Template-Mediated Synthesis and Characterization of Lariat RNA

Encouraged and intrigued by the successful ligation of the DNA lariat **3.16**, we were tempted to undertake RNA lariat synthesis using the same bimolecular complex design. Both DNA lariats synthesized thus far were only topologically related to the naturally occurring splicing intermediate, so fundamentally, only a lariat oligonucleotide containing both oligoribonucleotide loop and stem motifs would share biochemical similarities with the true intronic structure. The intermolecular template-mediated chemical ligation model potentially allows one to introduce any base composition within the loop and stem structure of the lariat, without being limited to those precursors that contain internal sequence complementarity or rely on purine-rich templates for ligation *via* a triple helical complex. The branched RNA lariat antecedent (**3.20**; **Table 3.5**) was constructed using our well-established protocols for the convergent synthesis of branched RNAs.[204,206,230,233] Maintaining the natural intronic structure, the RNA lariat precursor embodied a riboadenosine branchpoint, with the 5'-, 3'- and 2'-branch junctions consisting of oligoribonucleotide appendages rather than deoxynucleotides as in the DNA lariat. The branching synthon utilized in this case was an adenosine 2',3'-O-bis-phosphoramidite reagent **1.1**, which was synthesized as described previously. Furthermore, the sequence design incorporated a G-nucleotide at the 2',5' and 3',5' positions of the branch in order to permit efficient characterization of the lariat structure by hydrolysis with the RNA lariat debranching enzyme, a specific 2',5'-phosphodiesterase.[109,112,115,208,212] Since two neighboring strands are tethered together *via* the branching agent, the oligonucleotide sequences found at the 2' and 3' positions of the branch junction are identical in base composition. The complementary regions in the 2'- and 5'-wings mirrored those sequences used in the DNA lariat study. As such, and the 2'- and 5'-branch extensions would be able to associate with a complementary template splint so as to place the 3'-phosphate/5'-hydroxyl junction in the correct spatial geometry for effective ligation. The optimal reaction constraints utilized for the synthesis of the DNA lariat (**3.16**) were maintained in this study, such as a 14-nt template bound to a 7 + 7 nucleotide complementary region. Furthermore, the favored secondary structure of the loop region was calculated using the RNA MFOLD program,

however this time, reaction constraints such as temperature and ionic strength of the ligation buffer could not be inputted as these parameters were already defined (*i.e.* 1 M NaCl, T=37°C). Once again, computer analysis established that the 37-nt loop structure would likely fold into one probable hairpin-like motif (ΔG=-0.7 kcal/mol), stabilized by two Watson-Crick interactions between C_6/G_{33} and G_7/C_{32} together with a wobble contact between U_8/G_{31} (**Figure 3.18**). In addition, to ensure isomeric 3',5'-phosphodiester bond purity during the ligation reaction, deoxythymidine residues were integrated at both the 5' and 3' terminal positions. Besides the very productive contact that these two pyrimidines make at the nicked junction, this was done to assure chemical purity at the ligation site, since the nonenzymatic ligation may otherwise result in mixtures of 2',5' and 3',5' diesters when ribonucleotides are joined.[249,305,310,323]

Table 3.5: Sequences of RNA oligonucleotides used in the intermolecular template-mediated synthesis of an RNA lariat.

Code	Sequence (5'→ 3')	Designation
3.18	GCGUUGt$_P$ ^{HO}tGUUGCG	**Nicked 3'- & 5'-RNA Control**
3.19	GCGUUGttGUUGCG	**Full Length RNA Control**
3.20	^{HO}tGUUGCGU$_{11}$A$^{2'5'}$(GU$_{10}$GCGUUGt$_P$) $_{3'5'}$GU$_{10}$GCGUUGt$_P$	**Lariat-RNA Precursor**
3.21	CGCAACAACAACGC	**RNA Template**

Notation: t=deoxythymidine nucleotide; large caps=ribonucleotide residues; OH=terminal hydroxyl; P=terminal phosphate; underlined oligonucleotides represent those which are complementary to the intermolecular DNA and RNA templates **3.17** and **3.21** respectively.

Briefly, the forked structure was assembled in the normal 3'→5' direction on a medium-density deoxythymidine functionalized support (60 μmol/g) in order to ensure maximal coupling between adjacent strands once the bis-phosphoramidite was introduced. Although lower nucleoside loadings are typically favored for the synthesis of long oligonucleotides (>40-mers), maximizing branching efficiency using a higher CPG loading was key to ensuring a sufficient yield of lariat precursor. In order to compensate for this, a 1000 Å pore-sized support was used instead of the more common 500 Å. A larger pore diameter assures adequate distance between the strands, since smaller pores

tend to fill up with the growing oligonucleotide and steric hindrance severely reduces the efficiency of coupling further residues. CPG functionalized with any nucleoside could have been used at this point, since the first step in chain assembly involved the incorporation of the commercial phosphorylation reagent followed by strand elongation from the free hydroxyl of the sulfonylethoxy group (**Figure 3.6B**). As mentioned previously, ammonia deprotection results in the concomitant cleavage of the support bound nucleoside and the bond between phosphorylation reagent and the second nucleotide added, in this case a 2'-deoxythymidine, thereby releasing two contiguous 3'-phosphate termini, at both the 2',5' and 3',5' extensions from the branch.

Chain elongation utilized commercially available 2'-O-*tert*-butyldimethylsilyl phosphoramidites and standard coupling times, while the branching synthon necessitated a lower concentration (0.03 M) and extended coupling time (30 minutes) for sufficient reaction.[204,206] The fully deprotected oligomer (**3.20**) was then analyzed and purified by denaturing PAGE and its nucleotide composition characterized by MALDI-TOF-MS (Experimental **Table 7.3**). PAGE analysis of the crude lariat precursor indicated that the branching reaction proceeded with an approximate 50% yield (**Figure 3.24**). Evidently, failure of the branching synthon to react with the 5'-terminus of the support bound oligomer produced a small amount of truncated product (Lane 1; band c). The predominant abridged products were the linear isomeric failure sequences (37-nt) that result when the branching synthon fails to react with two adjacent strands (Lane 1; band b). Interestingly, **3.20** migrates through the gel as a slower rate than its corresponding full linear construct containing the identical sequence composition (compare lanes 1 & 2). Such an effect is likely due to the frictional influence of the branched topology as it migrates through the porous matrix. A sufficient yield of purified precursor **3.20** was isolated and employed in the determination of its thermodynamic behavior in the presence of complementary splint as well as its ligation potential.

RNA lariat precursor **3.20** was annealed to either its complementary template DNA (**3.17**) or RNA (**3.21**) at a 1:1 stoichiometric ratio in ligation buffer (0.25 M MES). Complexes were left to associate overnight at 4°C, and their thermal denaturation profiles monitored

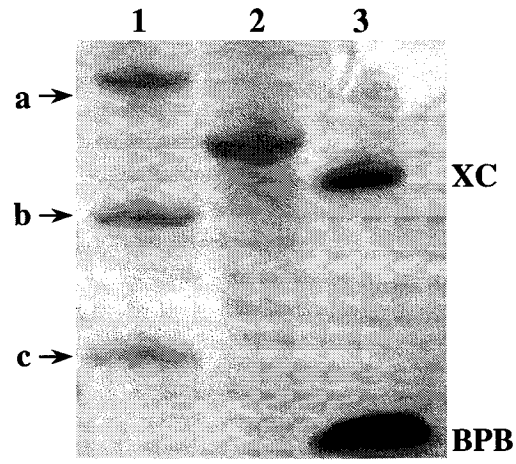


Figure 3.24: PAGE analysis (12% acrylamide; 8.3 M urea) demonstrating the convergent synthesis of **3.20**. Lane 1: crude synthesis of **3.20**; (a) full branched Y-RNA (55-mer) product **3.20**; (b) isomeric linear failure sequences (37-mers); 5'-tGUUGCGU$_{11}$A$_{3'5'}$GU$_{10}$GCGUUGt$_P$-3' and 5'-tGUUGCGU$_{11}$A$^{2'5'}$GU$_{10}$GCGUUGt$_P$-3'; (c) linear branching failure (18-mer); 5'-GU$_{10}$GCGUUGt$_P$-3'; Lane 2: linear form of **3.20** (unbranched) containing identical base composition: 5'-GU$_{10}$GCGUUGttGUUGCGU$_{11}$AGU$_{10}$GCGUUGt$_P$-3'; Lane 3: gel loading dyes XC (xylene cyanol) and BPB (bromophenol blue).

at 260 nm (**Figure 3.25**). Under the condensation conditions studied, nicked **3.20** formed a much more stable hybrid complex with the RNA splint than its corresponding DNA template (ΔT_m=22°C; **Table 3.6**). Suitably, a tandem oligonucleotide (**3.18**) proportionate to merely the complementary portions of **3.20** was synthesized, and its melting behavior in the presence of both splints monitored for comparative purposes (**Table 3.6**; **Figure 3.25**). A parallel trend was revealed in this case, whereby the RNA guide sequence assembles into a more thermally stable aggregate with the tandem RNA repeat than its DNA correlative (ΔT_m=16°C; **Table 3.6**). In general, RNA-RNA hybrid duplexes with mixed base composition tend to be more energetically stabilized than DNA-RNA hybrids of identical sequence.[19,324] The more rigid nature of the RNA backbone, composed of C3'-*endo* "compact" nucleotide repeats, presumably pre-organizes the structure for favorable and specific binding to its complementary target.[298,321] Specifically, a noticeable decrease in melting temperature is manifest when the DNA splint is hybridized to the nicked RNA lariat compared to its complex with the tandem RNA (ΔT_m=6°C; **Figure 3.25**), however, this is not observed with the

RNA template as the T_m for both complexes is identical. Even so, the helix-to-coil transition in the nicked loop structure is less cooperative than in the tandem RNA suggesting that full complementarity within the binding domain was not achieved (**Figure 3.25**). Consistent with the data obtained for the DNA lariat, the drop in overall complex stability in the DNA-RNA hybrid and the non-cooperative denaturation of the RNA hybrid seemingly stems from the secondary structure adopted by the loop portion of the lariat precursor and the difficulty imposed by accommodating the target stand in the stem-loop region (see Section 3.3.2). At lower ionic strength (10 mM NaCl), stable bimolecular or termolecular complexes of either the tandem RNA (**3.18**) or lariat precursor (**3.20**) with splint DNA were indeterminate (**Table 3.6**). A cooperative transition with a T_m of 23°C was observed for the tandem RNA complexed to its RNA template, yet its thermal stability was at least half of that observed in ligation buffer. Since the T_m's of all complexes in MES buffer were deemed to be sufficiently stable, we proceeded with investigating their ligation ability.

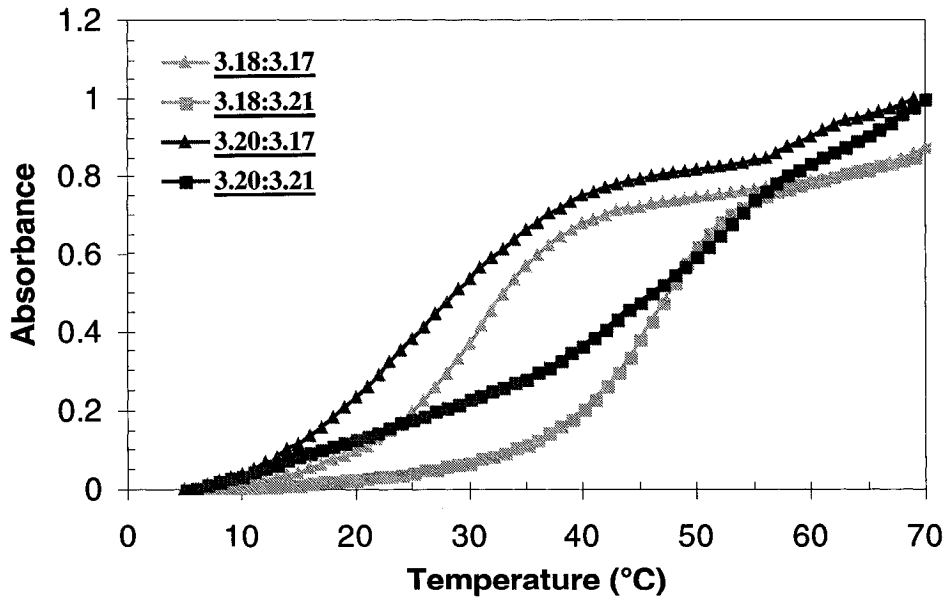


Figure 3.25: Comparative T_m analysis of nicked Y-RNA **3.20** and nicked template RNA **3.18** hybridized to their complementary DNA splint **3.17** or RNA splint **3.21**. Analyses were conducted under CNBr ligation conditions (Buffer: 0.25 M MES pH 7.6, 20 mM $MgCl_2$).

Table 3.6: Thermal denaturation and ligation data for nicked and ligated intermolecular RNA complements with their corresponding DNA and RNA template splints

Template	Complement	T_m (°C)		Ligation Yield (%)
		Tris[a]	MES[b]	
3.18	3.17	<15	31.0	
	3.21	23.0	47.1	
3.20-nicked	3.17	<10	25.0	
	3.21	-	47.0	
3.19	3.17	44.0	-	
	3.21	56.0	-	
3.20-ligated	3.17	-	-	46
	3.21	47.0	-	35

T_m values are the average of 3 successive runs and are within ±0.5°C. [a]10 mM Tris pH 7.5, 10 mM NaCl; [b]0.25 M MES pH 7.6, 20 mM $MgCl_2$. The ligation yield was determined by densitometric quantitation of the circularized product band using the UN-SCAN-IT (Silk Scientific) software program. dash=not determined.

The RNA lariat precursor **3.20** was mixed with an equimolar amount of either DNA splint (**3.17**) or RNA template (**3.21**) in ligation buffer at a total strand concentration of 10^{-4} M. Samples were heated briefly and allowed to cool slowly to the desired reaction temperature (4°C). A major disadvantage of using the convergent strategy for bRNA synthesis is that it introduces identical sequences at both the 2'- and 3'-extensions of the branch. During strand association, the complementary splint could potentially hybridize and align either or both the 5'- and 2'-extensions or the 5'- and 3'-extensions (**Figure 3.26**). This would fold the loop into two probable structures, leading to the formation of the desired 2'-lariat and unsought 3'-lariat RNA regioisomer following phosphodiester bond condensation. Given that previous data has demonstrated a preferred stacking interaction between the branchpoint adenine and the base at the 2'-position[314,315], we imagined that this factor may influence the ratio of folded 2'-loop structure leading to predominant ligation of the desired lariat. Analysis of the CNBr-mediated reaction products once again unveiled the presence of a distinct, faster moving oligonucleotide species, with a migration rate consistent to that of the DNA lariat described previously (**Figure 3.27**). Such a lariat would contain a 37-nt loop and an 18-nt stem. A negligible

amount of dimeric material, with an approximate molecular weight of 35 KDa (*i.e.* 110-nts) was also observed with both the DNA and RNA splints, formed by the termolecular association of two molecules of precursor with one copy of template. Densitometric quantitation of the resultant product bands indicated that the DNA guide was more effective at positioning the phosphate-hydroxyl junction for condensation, since the yield obtained was roughly 10% higher than with the RNA template (**Figure 3.27**; lanes 2 & 3). This was surprising since the thermal melting profiles demonstrated that the RNA template formed the most stable hybrid duplex with the precursor under the ligation conditions. As such, in this context it does not appear duplex stability is the dominating factor influencing ligation efficiency, and instead definition of the local stereogeometry at the nicked junction is of utmost importance. Unequivocally, duplex RNA exists as an A-form helix, and it is well established that the nucleotide repeats adopt a C3'-*endo* sugar pucker.[1] Since the ligation junction contains two flanking deoxythymidine residues, RNA template binding forces the deoxyribose sugars to adopt an intermediary

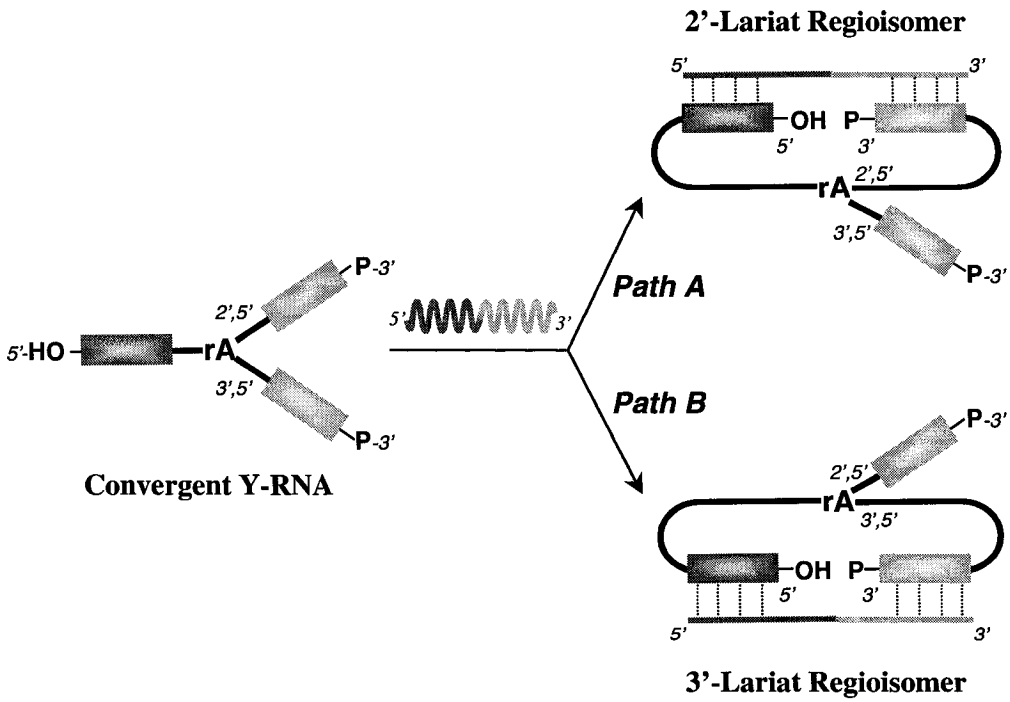


Figure 3.26: Convergently synthesized bRNA can assemble into two lariat precursor structures with complementary template giving rise to the desired 2'-lariat regioisomer and the unsought 3'-lariat regioisomer.

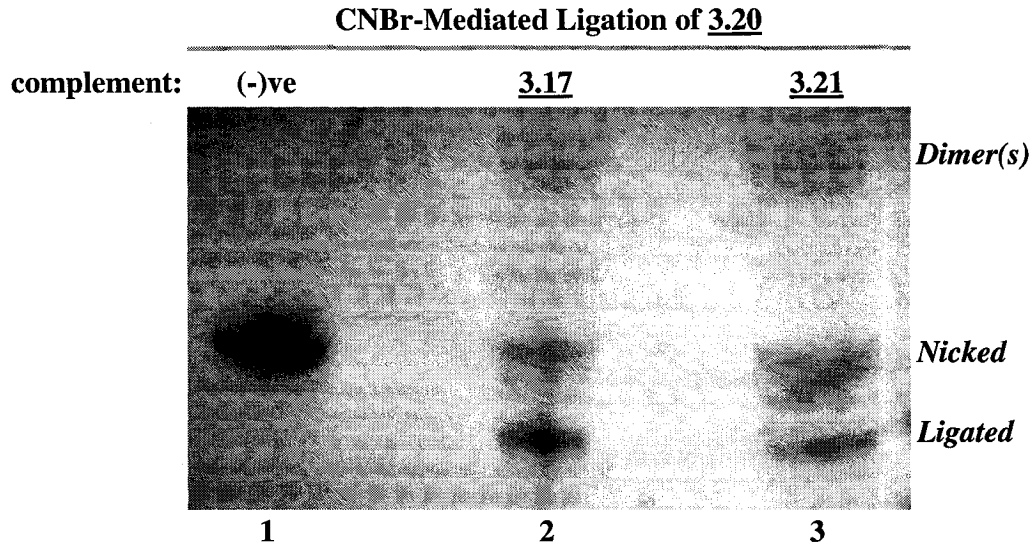


Figure 3.27: Analysis of the CNBr-mediated ligation of **3:20** in the presence of 1:1 complementary DNA splint **3.17** or RNA splint **3.21** by 12% denaturing PAGE (8.3 M urea). Lane 1: negative control **3.20** (no complementary splint); Lane 2: CNBr ligation of **3.20** in the presence of DNA splint 3.17; Lane 3: CNBr ligation of **3.20** in the presence of RNA splint **3.21**. All ligation reactions contained 10^{-4} M precursor and complementary splint dissolved in 0.25 M MES pH 7.6, 20 mM $MgCl_2$ buffer.

conformation between the normal DNA C2'-*endo* pucker and the C3'-*endo* conformation of the RNA-RNA duplex.[18] In this case, the 3'-phosphate would be placed in a pseudoequatorial position, thereby constraining it into a more sterically hindered orientation for interaction with the 5'-hydroxyl of the neighboring thymidine. Alternatively, when the DNA splint binds, the favorable C2'-*endo* local geometry is maintained, with some minor perturbations imposed by the adjacent ribonucleotide residues.[325] Essentially, this places the phosphate group in a pseudoaxial position, for constructive in-line attack by the adjacent 5'-hydroxyl moiety.[310] Even though the DNA splint aligns the reactive interface into the proper spatial orientation, the cyclization yields obtained were still only moderately acceptable (**Table 3.6**). In contrast to the DNA lariat, the phosphate terminus during bRNA synthesis was introduced as a first step, rather than at the end of the synthesis cycle, so theoretically, 100% of the 3'-termini should have been phosphorylated. Quite possibly, water molecules were consuming the activated phosphate groups and excess cyanogen bromide faster than they could react. To verify

this, nicked **3.20** was ligated in the presence of complementary DNA splint **3.17** under standard condensation conditions. After the usual five-minute incubation period at 4°C, a small aliquot of ligated material was removed from the mixture, and the remaining RNA complex treated with fresh CNBr for an additional five minutes. Analysis of both reaction mixtures demonstrated that the first incubation produced approximately 44% of ligated material, whereas the reaction with supplemental CNBr exhibited an appreciable intensification of lariat product (*ca.* 58%). Clearly, this suggests that hydrolysis of the activated phosphate is the rationale behind the lower than anticipated yields (data not shown). This technique has successfully been used to intensify the amount of ligation product produced even in strictly unfavored duplexes.[310] Slight changes in orientation and proximity of the reacting groups caused by a variation in sugar pucker decreases the coupling yield, which strongly supports the notion that the result of chemical ligation depends mainly on the reactive site structure, and the reaction effectiveness dictates how much the conformation of the juxtaposed groups differs from the canonical.

The anticipated RNA lariat product band was excised from a denaturing preparatory gel, extracted into water and desalted by SEC. The oligonucleotide composition was determined by MALDI-TOF-MS and yet again, the difference in molecular weight between the nicked and ligated forms of **3.20** was consistent with the condensation of the phosphate hydroxyl impingement. Furthermore, thermal melting analysis of the closed lariat **3.20** hybridized to its complementary RNA splint **3.21**, displayed enhanced thermal stability compared to its nicked tandem RNA **3.18** in 10 mM NaCl (**Table 3.6**; **Figure 3.28**). The resultant T_m was also 9°C lower than that of the full-length linear complementary region (**3.19**) complexed to the RNA splint, consistent with the results obtained for the DNA lariat. As described previously, this suggests that complementary RNA has a difficult time integrating itself within the circularized RNA loop structure for productive complex formation, resulting in a reduction in the overall hybrid stability.

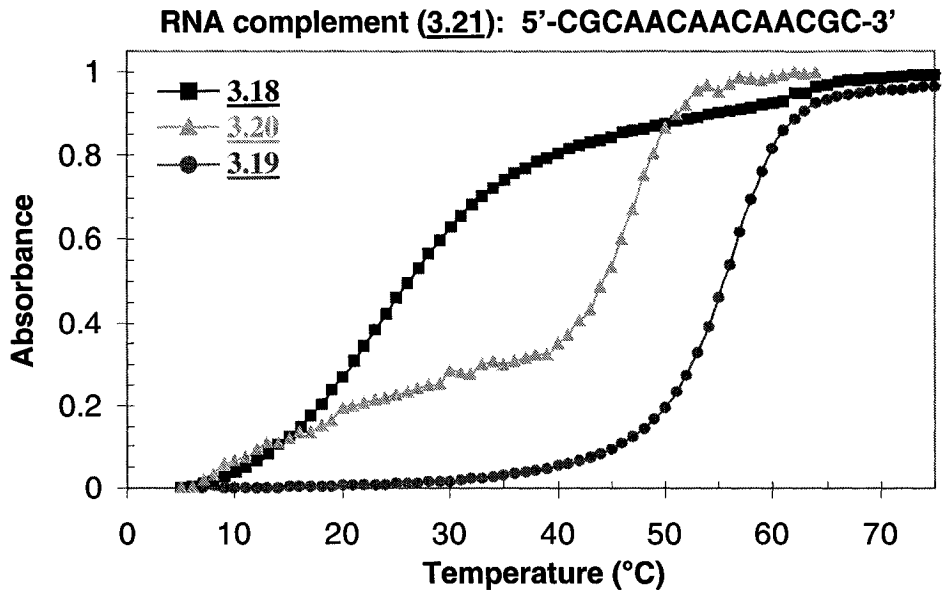


Figure 3.28: Characterization of ligated **3.20** by T_m analysis and comparison to the nicked (**3.18**) and full (**3.19**) template regions of **3.20** by hybridization to the complementary RNA splint **3.21**. Analyses were conducted in 10 mM Tris-HCl pH 7.5, 10 mM NaCl buffer.

3.3.5. Regioselectivity of RNA Lariat Cyclization

Encouraged by the presence of only one new product band, concurring with the anomalous migratory rate of a circularized lasso species, we were inclined to believe that only one lariat regioisomer, either the 2'-cyclized or 3'-cyclized, had prevailed. Although the two regioisomeric species would embody indistinguishable molecular weights, it seemed feasible that they might adopt diverse topological structures, thereby permitting their resolution on a more highly crosslinked gel matrix (*i.e.* 24 % acrylamide).[93,94,101] However, even under less porous denaturing PAGE conditions, the lariat product band did not separate into two distinct species. The same was true when the cyclization product was analyzed under non-denaturing gel conditions and by both anion-exchange and reverse-phase HPLC (data not shown). In order to further characterize the lariat structure and determine if one regioisomer indeed predominated over another, an RNA debranching assay was performed. Debranching of the lariat product should proceed

efficiently since the synthetic structure incorporated a wild-type G-nucleotide at the 2'-position of the adenosine. Purine-containing branches have been shown to be the preferred substrates for hydrolysis.[115,116,208,212] The individual regioisomeric ligation products could be easily distinguished using this assay, since cleavage would result in completely different hydrolysis products (**Figure 3.29**). Specific hydrolysis of the desired 2'-lariat ligation product at its 2',5'-phosphodiester bond would convert it to a linearized RNA molecule consisting of 55 nucleotides (**Figure 3.29A**). On the other hand, treatment of the undesired 3'-lariat regioisomer with DBR would produce two digestion products: an 18 nucleotide product derived from the stem of the molecule and a 37 nucleotide circular RNA containing the riboadenosine branchsite (**Figure 3.29B**).

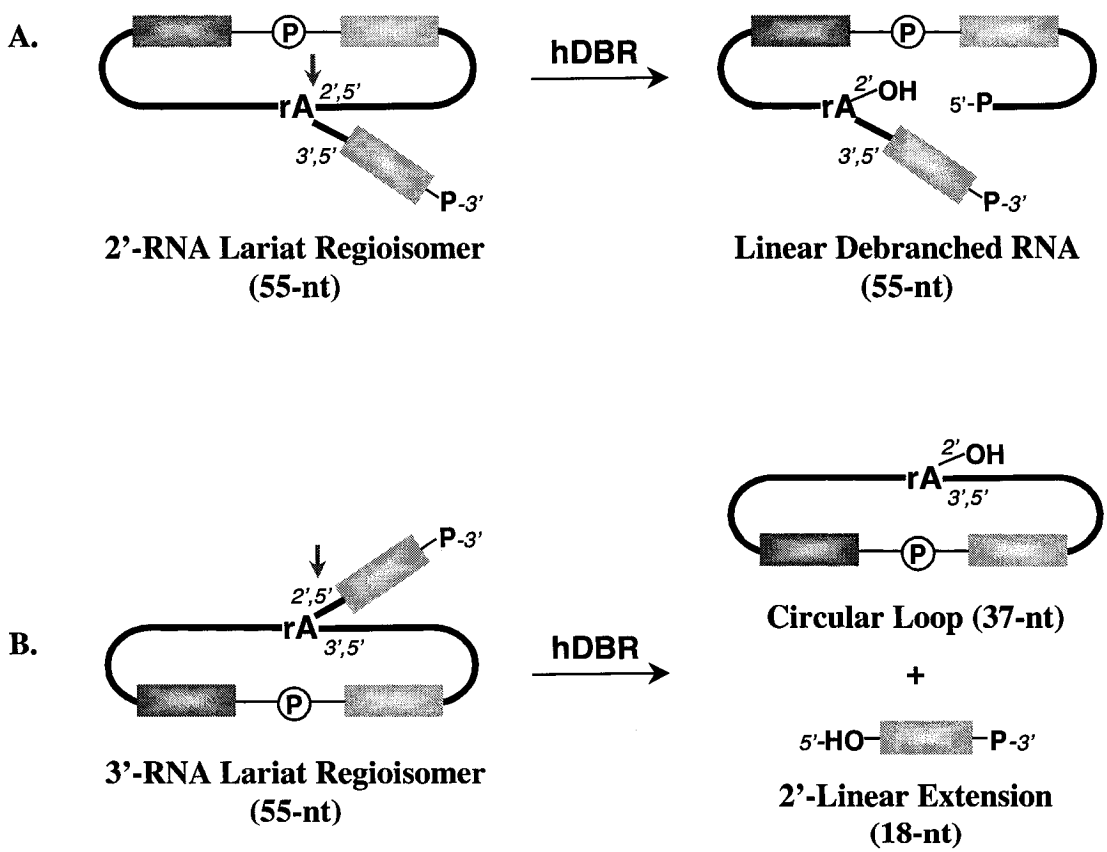


Figure 3.29: Schematic representation demonstrating hydrolysis of regioisomeric 2'- and 3'-lariat forms of **3.20** using the HeLa debranching enzyme (hDBR) found in nuclear extract. Red arrow represents the DBR hydrolysis site. Panel A: Debranching of the 2'-lariat regioisomer; Panel B: Debranching of the 3'-lariat regioisomer.

Given that both regioisomers were conceivably present, hydrolysis of a 3'-terminally radiolableled substrate mixture, followed by PAGE analysis would reveal only two bands; the linearized 55-mer and 18-nt strands. The circular 37-mer would be concealed since it does not bear a radioactively labeled 3'-end. Substrates could not be radiolabeled using T4 polynucleotide kinase since they did not bear any free 5'-ends. Prior to radioactive probe incorporation, the substrate lariat was dephosphorylated at the 3'-terminus with calf intestinal alkaline phosphatase (CIAP). Attempts to radiolabel the lariat at the 3'-terminus using two different enzymatic methods failed miserably. Firstly, 3'-end labeling of the thymidine terminus with the deoxyribonucleotide specific enzyme terminal deoxynucleotidyl transferase (TdT) and α-[^{32}P]-ddATP resulted in an insignificant amount of radiolabel incorporation (data not shown). It was later discovered that although TdT is highly effective at adding deoxynucleotides to the 3'-terminus of both single and double-stranded DNA, its minimal substrate requirement is a three-unit stretch of oligodeoxynucleotides.[326] Alternatively, efforts were made to radiolabel the terminus with T4 RNA ligase and α-[^{32}P]-cytosine-3',5'-biphosphate.[327,328] Although deoxyribonucleotides are substrates for the enzyme[263,327,329,330], albeit poorer than ribonucleotides, only a very minute amount of labeled material was isolable. The likely problem appeared to be incomplete dephosphorylation of the 3'-terminus with calf intestinal alkaline phosphatase (CIAP) prior to end-labeling. Furthermore, the labeled material that was isolated turned out to be completely degraded. In the end, cold, substrate lariat was used and the hydrolysis products visualized by staining of the gel with the commercially available Stains-All®. Once again, we exploited the debranching activity found in HeLa nuclear extract (hDBR).[109] EDTA (10 mM) was added to the assay mix as a general exonuclease inhibitor.[109] Debranching of the purified lariat **3.20** resulted in the production of three new product species, two of which co-migrated with appropriate oligonucleotide standards (**Figure 3.30**). The most retained product band coincided with the full length linear 55-nt RNA; the debranching product of the desired 2'-lariat regioisomer (Lanes 2 & 3). On the other hand, another product band co-migrated with the 18-nt linear control; one of the two debranching products of the 3'-lariat isomer

(Lanes 2 & 4). An additional band, which migrated slightly faster than the linear 55-nt product was also evident (Lanes 2 & 4). Provided that both regioisomeric lariats were present in equivalent amounts, as determined by densitometry, this unknown oligonucleotide band was attributed to the circular 37-nt loop structure. Indeed, it appeared that cyclization of the 2'-branch extension and the 3'-appendage occurs with equal efficiency, and it is likely that the base-stacking interaction between the adenine base at the branchpoint and the 2',5'-linked guanine is too weak to drive circularization of one extension over the other in such a large loop structure.

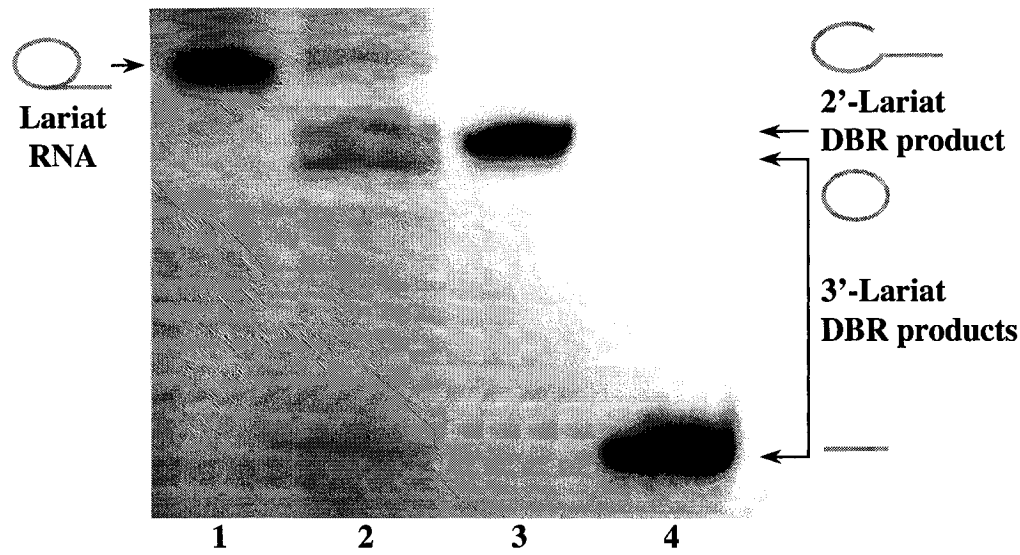


Figure 3.30: 24% denaturing PAGE (7 M urea) demonstrating the debranching of ligated **3.20**. Lane 1: ligated lariat RNA **3.20**; Lane 2: ligated **3.20** + 30% HeLa extract (debranching); Lane 3: 2'-lariat debranching product 55-mer (5'-$^{P}GU_{10}GCGUUGttGUUGCGU_{11}AGU_{10}GCGUUGt_{P}$-3'); Lane 4: 3'-lariat debranching product 18-mer (5'-$^{P}GU_{10}GCGUUGt_{P}$-3').

At the same time that this thesis was being compiled, a report emerged concerning a *de novo* approach to synthesizing lariat RNA based on a deoxyribozymes isolated from random DNA pools that contained RNA ligase activity (*i.e.* DNA-dependent RNA ligase).[331,332] The deoxyribozymes selected created non-native 2',5'-phosphodiester bonds from a linear RNA species with greater than 95% selectivity for the non-native bonds versus the native 3',5'-linkage. Nonetheless, the precursor RNA substrates were synthesized *via* enzymatic T7 RNA polymerase transcription from a DNA template, and

as such, yields of precursor are limited in this scope. Furthermore, the apparent yields of lariats observed, depending on the deoxyribozyme species used, were only in the 1-15% range. Our method provides a significant advantage since the precursor substrates are synthesized chemically, *via* well-established solid-phase methodologies, thereby allowing us to isolate an appreciable amount of precursor for larger scale-production. In addition, the template-directed cyclization is clearly more efficient, since the amount of circularized lariat product observed in all cases was in the 40-60% range. Regardless, the results from the deoxyribozyme-mediated ligation are indeed of considerable mechanistic interest, given that nature's spliceosome is likely a ribozyme in itself.[80,81,83]

3.4. CONCLUSIONS

Overall, we have successfully demonstrated that lariat oligonucleotides of both DNA and RNA circular loop and stem structures can be synthesized using template-mediated methodologies. The results reported herein validate the first examples of the synthesis of medium-sized DNA and RNA lariat oligonucleotides using well-established solid-phase phosphoramidite chemistry and commercially available reagents. Both the intramolecular dumbbell strategy and the intermolecular splint/complement strategy are effective, however, only the latter allows for the introduction of any base composition within the loop and stem structures and is not limited to those molecules containing internal templates. Furthermore, it permitted the synthesis of a biologically relevant RNA lariat molecule containing the wild type adenosine branchpoint and guanosine nucleotide at the 2'-appendage of the branch, which was a substrate for the human lariat debranching enzyme.

More specifically, we have shown in the case of the dumbbell structures, that the insertion of a normally destabilizing 2'-rA insert is well tolerated in the complex state owing to the stabilizing effects of the T_4 homonucleotide loop motif. Analysis of the ligation properties of various circular dumbbells allowed us to predict the most favorable sequence arrangement for successful ligation of the branched dumbbell precursor.

Particularly important factors influencing the nick-sealing efficiency were the use of a 5'-hydroxyl/3'-phosphate abutment at the nick site, the insertion of a productive 5'-T/3'-pT ligation junction, as well as shifting the ligation site to a more distal position from the locally distorted 2'-rA insert. Ligations could be performed in a matter of minutes using cyanogen bromide as a condensing reagent. Nonetheless, the relative yield of condensed lariat product **3.12** was lower than anticipated, compared to its circular counterpart **3.10** owing to incomplete phosphorylation of the 3'-terminus of the molecule. Thermal denaturation analysis (T_m) and circular dichroism (CD) of the nicked and ligated forms of the DNA dumbbell lariat **3.12** also revealed that the 3'-hexadecanucleotide extension from the branchpoint does not alter the overall helix geometry of the loop structure and likely projects out into the solvent in accordance with a base-stacking interaction between the adenine base at the branchpoint and the 2',5'-linked cytidine. In addition, the lariat nature of the molecule was confirmed by spectroscopic (T_m) and enzymatic (BSPDE) techniques, which indicated that the loop motif was indeed circularized.

In the case of the bimolecular synthetic approach to lariats, effective ligation was eminent in the presence of a 1:1 ratio of template to complementary branched precursor oligonucleotide using CNBr as a condensing reagent. The resultant T_m's of the nicked DNA (**3.16**) or RNA (**3.20**) branched precursors in the presence of either a DNA splint (**3.17**) or RNA splint (**3.21**) were lower than their respective complementary regions only (DNA: **3.13**, **3.14**; RNA: **3.18**), suggesting that the nicked precursor was folded into the desired hairpin-like structure, and as such was incapable of comfortably accommodating the complementary template strands in the kinked portions of the stem-loop region. The same rationale appeared to hold true in the case of the ligated lariat structures, whereby the T_m of the closed complexes with their corresponding DNA and/or RNA splints were lower than for the full complementary regions (**3.15** and **3.19**) reminiscent of the hybridizing portion of the lariat. Nonetheless, a significant enhancement in T_m was observed between the nicked and ligated forms of the DNA and RNA lariat molecules, indicating that circularization of the loops was realized. Ligation of the branched DNA precursor resulted in a regioisomerically pure 2'-linked lariat molecule, given that the branched DNA synthetic methodology incorporated diverse sequences at the 2' and 3'

branch appendages. On the other hand, branched RNA synthesis integrated symmetrical sequences at the branch extensions, and as such, both the 2' and 3'-linked lariat regioisomers were formed in equal amounts. In order to prevent this, a synthetic strategy for synthesizing branched RNA with unlike extensions at the branch junction is imperative. Methods for constructing unsymmetrical branched RNA molecules have been investigated[218,226], however, the approaches utilize many non-standard nucleoside building blocks and are highly impractical for the synthesis of long branched RNA sequences required for biologically applicable lariat formation. We are currently investigating the possibility of separating the two lariat regioisomers *via* HPLC or capillary electrophoresis (CE).

CHAPTER 4: SELECTIVE AND POTENT INHIBITION OF HIV-1 REVERSE TRANSCRIPTASE (HIV-1 RT) BY AN RNA DUMBBELL

4.1. INTRODUCTION

The disease known as acquired immunodeficiency syndrome (AIDS) was first reported in an issue of *Morbidity and Mortality Weekly Report* (MMWR), a bulletin published by the Centers for Disease Control (CDC) on June 5th, 1981.[333] More than twenty years later[334-336], it is estimated that nearly 40 million individuals worldwide are living with AIDS, a fatal disease that overwhelms the body's immune response system.[337,338] At the present time, no known cure for this devastating illness is available, however, numerous research endeavors are attempting to attack the disease from various fronts. The root cause of the disease, an RNA retrovirus suitably termed the human immunodeficiency virus type 1 (HIV-1), was identified and isolated by a French research team in 1983.[339,340] Not long after, HIV-2, another strain of the virus was isolated and characterized,[341,342] yet it showed to be much less infectious than the former subtype and resulted in a slower onset of the disease.[343] After initial contact and attachment of the HIV virion to a host cell of the immune system (*e.g.* lymphocytes, monocytes), a cascade of intracellular events erupt (**Figure 4.1**). The consequence of these concomitant events is the production of massive numbers of new viral particles, death of the infected cells, and ultimate devastation of the immune system. The gravity of this disease and the epidemic rate at which it has spread have prompted the search for therapeutic agents which can suppress or eradicate it before the inception of significant immune system damage.

A single HIV virion begins its proliferative infection of a susceptible host cell by specific binding to an extrinsic CD4 receptor; an immunoglobin-like protein expressed on the surface of many lymphocytes, which are critical in defending the body's immune system. HIV enters the cell *via* receptor-mediated endocytosis, where the virion nucleocapsid proteins are uncoated and the genomic viral RNA material is expelled into the cytosol (**Figure 4.1**). At this point, the viral RNA is reverse-transcribed into a proviral

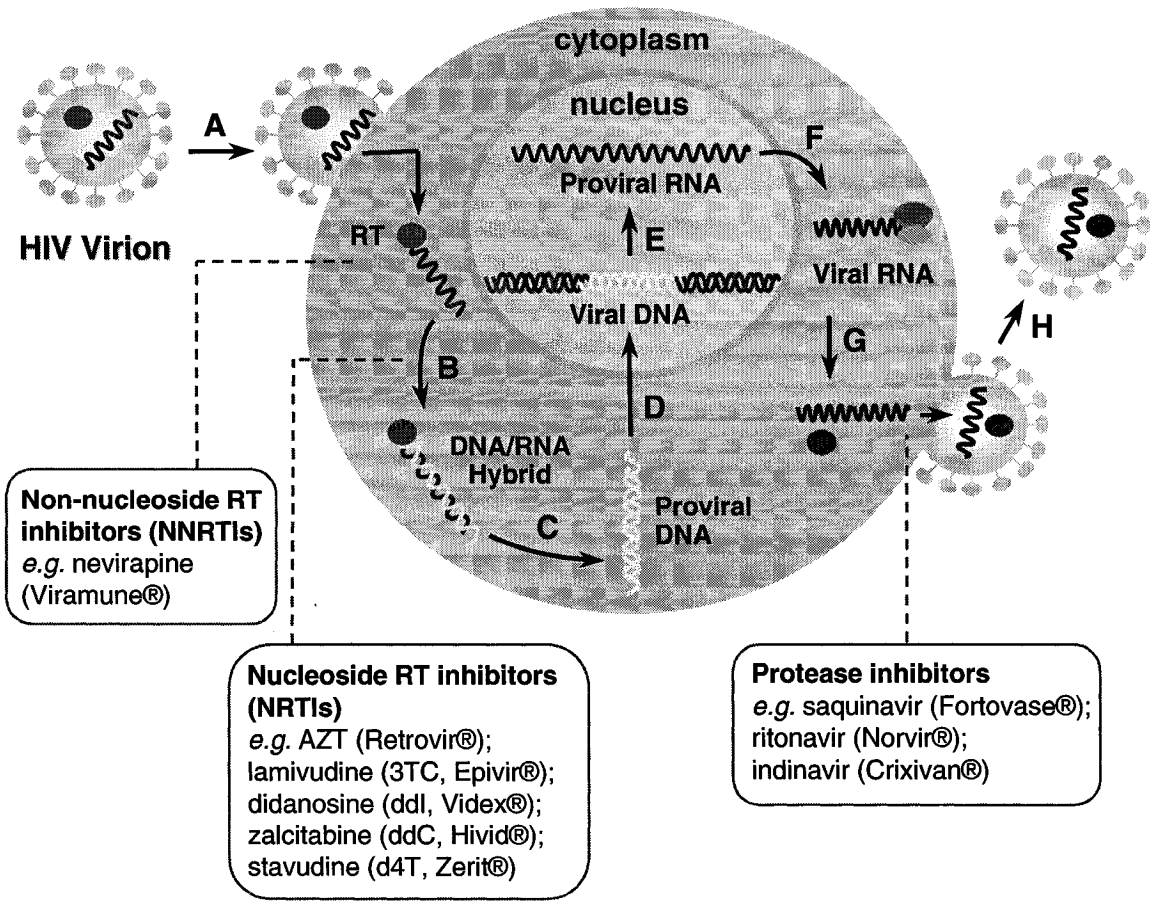


Figure 4.1: Schematic representation of the HIV replicative life cycle and the intervention stages of the three major classes of antiretroviral therapies. The key stages in viral progression are: (A) virus fusion with the host cell lymphocyte and release of viral genetic material; (B) Reverse transcription of viral RNA into complementary DNA (cDNA) by reverse transcriptase (RT); (C) RNase H mediated degradation of the viral RNA strand in the DNA/RNA hetroduplex and synthesis of proviral DNA (dsDNA) by RT; (D) Integration of proviral DNA into host cell genome by viral integrase; (E) Transcription of integrated DNA into proviral RNA; (F) Nuclear export of viral RNA; (G) Translation of viral mRNA into polyproteins for cleavage into functional viral proteins by viral protease; (H) Budding of mature virion particles from host cell.

complementary DNA (cDNA) by way of reverse transcriptase (RT), an enzyme which plays a cardinal role in the life cycle of the retrovirus.[344] This pluripotent enzyme contains three distinct enzymatic activities: an RNA-dependent DNA polymerase (RDDP) activity, a DNA-dependant DNA polymerase activity (DDDP) and a ribonuclease H (RNase H) activity. The viral RNA acts as a template, annealed to a $tRNA^{lys3}$ primer sequence at its primer binding site (PBS), thereby directing the

incorporation of deoxyribonucleotide triphosphates (dNTPs) by the RDDP. The resultant RNA/DNA heteroduplex is recognized by the RNase H domain of reverse transcriptase, and degrades the RNA portion of the resultant hybrid. The DDDP activity then synthesizes a new DNA second strand, yielding a proviral DNA duplex. This duplex DNA is transported across the nuclear membrane into the nucleus where a virally-encoded *integrase* enzyme inserts the proviral DNA into the host's genetic material. Once the viral DNA is integrated into the host's genome, it is possible that HIV may persist in a latent state for many years.[345] When the immune cell becomes activated, this latent provirus is stimulated and instructs the cellular machinery to produce the necessary components of HIV. Transcription is initiated, yielding viral RNAs that are processed and exported from the nucleus. One of the mRNA strands is translated into the protein assembly of HIV, which are cleaved by viral *protease* to give smaller protein subunits, whereas the other strand becomes the genetic material for the new viruses. As a final step, the newly manufactured virion particles congregate, are packaged in a new viral envelope, pinch off and enter into circulation, ready to continue infecting other host cells.

Antiretroviral therapeutics have been designed to intervene with any of the stages of the HIV replicative cycle described above, thereby suppressing viral proliferation. Of the myriad of lead compounds studied, only those that specifically target the retroviral activity of reverse transcriptase or the viral protease enzyme have achieved clinical success and have been approved for HIV therapy (**Figure 4.1**).[346,347] Combination therapy (protease + RT inhibitors) is now the standard care for patients with late-stage HIV, and the increasing trend is to prescribe synergistic treatments of antiretroviral agents for patients with less advanced forms of the disease.[348] Reverse transcriptase has been a prominent target in the discovery and development of antiviral chemotherapeutic strategies to suppress or prevent HIV infection owing to its multifunctional nature and critical role in viral replication.[344,349] In a mature HIV-1 virion, the reverse transcriptase enzyme exists as an asymmetric heterodimeric entity consisting of a 66 kDA (p66) and 51 kDA (p51) subunit (**Figure 4.2**).[350] The p51 subunit is produced by proteolytic cleavage of the p66 moiety and lacks the RNase H domain. On the other hand, both the polymerase and RNase H activities of RT are found in the p66 unit with

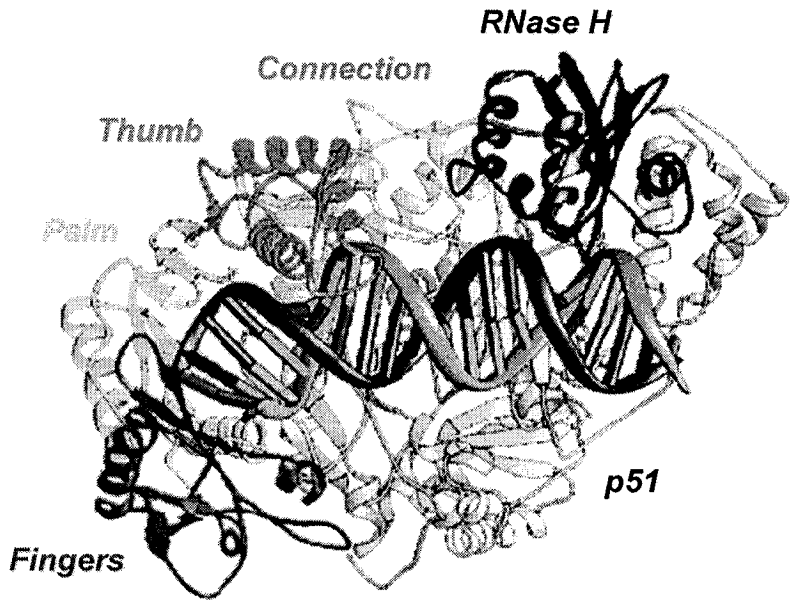


Figure 4.2: The X-ray crystal structure of the HIV-1 RT catalytic complex crosslinked with a DNA template/primer at a 3.2Å resolution (adapted from Huang, K *et. al. Science* **1998**, *282*, 1669-1675). A view of the RT catalytic complex with the polymerase active site on the left and the RNase H domain on the right. The domains of p66 are in color: fingers (red), palm (yellow), thumb (orange), connection (cyan), and RNase H (blue); p51 is in gray. The DNA template strand (light green) contains 25 nucleotides, and the primer strand (dark green), 21 nucleotides.

the former situated at the N-terminus and the latter in the C-terminus of the protein. Numerous crystal structures of the active HIV-1 RT complex have been obtained in an unliganded form[351] and in the presence of inhibitors[352,353] and nucleic acids.[354,355] The p66 domain resembles a human hand and is subdivided into five subdomains. The DNA polymerase activity is localized in four of those subdomains appropriately named the fingers, palm, thumb, and connection units (**Figure 4.2**). The latter links the activities of the DNA polymerase and the fifth subdomain, RNase H.[352,356,357] RNase H is known to require divalent metal ions for activity, with the highest activity typically seen with Mg^{2+} ions.[358] The first solution structure of the RNase H domain of HIV-1 RT was published this year and demonstrates that in the absence of nucleic acid substrate, two divalent magnesium ions (Mg^{2+}) can be well accommodated in the active site of the

enzyme.[359] Except for the active site pocket, the crystal structures of *E. coli* RNase H and HIV-1 RT RNase H are very similar despite their relatively low sequence homology.[360]

Several potent inhibitors have been developed that successfully antagonize the polymerase function of RT. In fact, all of the drugs currently approved as HIV-1 RT inhibitors are directed against its polymerase activity. Most of these therapeutics act by inhibiting RT globally rather than diminishing a particular step in the reverse transcription process. The two classes of drugs developed to block the polymerase activity of RT are the nucleoside analog RT-inhibitors (NRTIs) and the non-nucleoside RT-inhibitors (NNRTIs) (**Figure 4.1**). NRTIs are delivered into the system as prodrugs and compete with cellular dNTPs for incorporation into the growing viral DNA strand. Once phosphorylated by cellular kinases, the NRTIs are integrated into growing DNA strand, thereby inducing chain termination since the analogs lack a free 3'-hydroxyl group. Alternatively, the NNRTIs act as non-competitive antagonists of RT thereby preventing its normal retroviral function.[361] Unfortunately, the prolonged clinical use of each of these classes of compound has been shown to give rise to drug resistant mutations in the viral RNA and altered RT protein structure that confer resistance to these drugs.

A search for alternative treatments and more specific antiviral agents has lead to a relatively new class of RT inhibitors termed the oligonucleotide reverse transcriptase inhibitors (ONRTIs). A phosphorothioate (PS) oligonucleotide [$Sd(C)_{28}$] was found to competitively inhibit HIV template-primer binding to the RT enzyme with exceptionally high affinity, thereby inhibiting DNA synthesis.[362] Binding appeared to be non-specific for the HIV-1 DNA polymerase activity and as such toxicity may be an intrinsic problem with such phosphorothioate constructs. Similarly, a PS-oligonucleotide derived from the primer $tRNA^{lys3}$ sequence demonstrated strong inhibition of HIV-1 RT with an IC_{50} in the 40 nM range.[363] Potent competitive inhibition profiles have also been obtained with a series of RT oligonucleotide inhibitors based on RNA pseudoknots.[364] These "aptamers" were distinctively selected from a randomized pool of RNA molecules

by high-affinity binding to the target protein using the SELEX technique (**S**ystematic **E**volution of **L**igands by **Ex**ponential Enrichment).[365] Nonetheless, most of the oligonucleotide inhibitors shown to inhibit RT activity do so by competitively inhibiting the polymerase function of the enzyme, without specificity for its RNase H domain.

4.2. Antiretroviral Targeting of the HIV-1 Reverse Transcriptase RNase H Functionality

The RNase H activity of HIV-1 RT is vital for viral replication since it is specifically required to cleave the RNA portion of a DNA/RNA heteroduplex intermediate, thereby permitting the viral DNA to disengage and invade the host cell's genetic material (**Figure 4.3**). Furthermore, point mutations in the RNase H domain of RT provoked a marked decrease in the level of virus proliferation, demonstrating that a functional RNase H activity is essential for retroviral replication.[366] As such, RNase H is a highly attractive target for the development of new antiretroviral agents. There are several published examples of HIV-1 RNase H inhibitors *in vitro*, however, it is still unclear whether their effect on this enzymatic function is mediated by direct binding to the RNase H domain.[367] The blocking of the RT-associated RNase H activity has mostly been demonstrated in cell-free systems. For example, the RNase H activity of RT can be inhibited by 3'-azidothymidylate 5'-monophosphate (AZT-MP), a major intracellular metabolite of the NNRT inhibitor AZT with an IC_{50} in the 50 μM range. [368,369] Apart from a high inhibitory concentration requirement, the activity of AZT-MP was also dependent on metal cation identity, with Mg^{2+} as the most effective co-activator. Alternatively, a metal chelator *N*-(4-*tert*-butylbenzoyl)-2-hydroxy-1-naphthaldehyde hydrazone (BBNH) has demonstrated potent inhibition of RNase H activity (IC_{50}=3.5 μM), and is effective against mutant RT's that have high-level resistance to other NNRTIs.[370] Nonetheless, the compound acts by effectively inhibiting the DNA polymerase activity as well, likely resulting from interaction of BBNH with at least two different sites on the enzyme, including the RNase H catalytic site. Illimaquinone, a

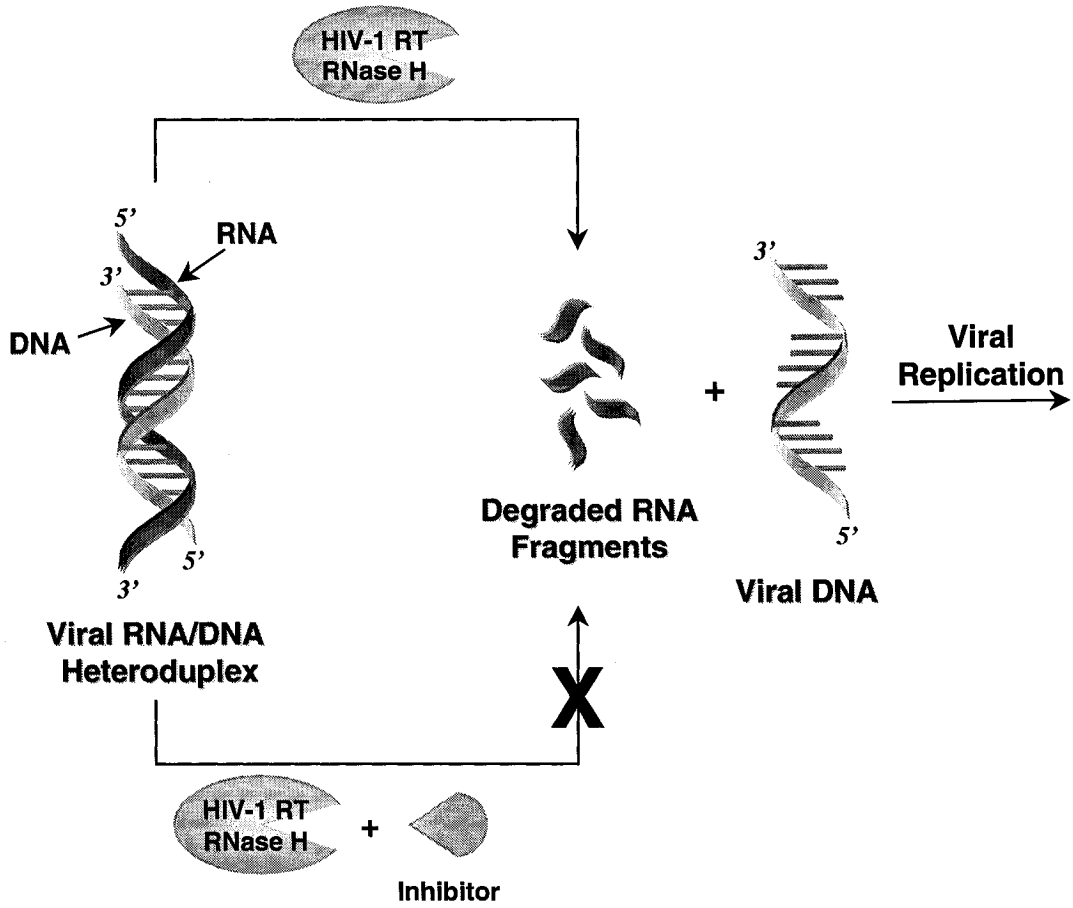


Figure 4.3: HIV-1 RT RNase H-mediated destruction of the viral RNA template strand in an RNA/DNA heteroduplex. In the absence of an RNase H inhibitor, HIV-1 RT degrades the RNA strand into small RNA fragments and releases the viral DNA copy which can go on and be integrated into the host cell's genome. Alternatively, specific inhibiton of RNase H activity leads to RNA/DNA hybrid arrest, and prevents HIV replication.

natural product of marine origin, inhibits the RNase H activity of HIV-1 RT preferentially, nevertheless, it also hinders the RNase H functionality of HIV-2 RT, MLV-RT and *E. coli*.[371,372] Most recently, a novel diketo acid compound: 4-[5-(benzoylamino)thien-2-yl]-2,4-dioxobutanoic acid has been shown to selectively bind and inhibit the RNase H domain with an IC_{50}=3.2 μM with little or no effect on DNA polymerization (IC_{50}>50 μM).[373]

Very few oligonucleotide based inhibitors (ONRTIs) specific to the HIV-1 RT RNase H activity have been reported to date. Potentially, such constructs can act by blocking the

enzyme's active site or impeding the binding of the viral DNA/RNA heteroduplex to the RNase H domain (**Figure 4.3**). Phosphorothioate oligonucleotides have demonstrated RNase H inhibition, however they also affect the DNA polymerase activity as mentioned previously.[374] Most recently, a series of DNA aptamers with high affinity and specificity for the RNase H activity of HIV-1 RT were isolated by SELEX. The most potent inhibitors were based on G-quartet motifs with IC_{50} values in the 500 nM range, yet they similarly inhibited the DNA polymerase function.[375] RNA aptamers have also displayed an equivalent dual inhibitory capacity.[376] Duplexes consisting of 2',5'-RNA/RNA were also shown to competitively suppress binding of the viral DNA/RNA substrate to HIV-1 RT without evoking its RNase H activity.[290] The effect on the polymerase activity was not studied. In a similar fashion, RNA duplexes have also demonstrated RNase H inhibitory activity of HIV-1 RT through competitive inhibition effects.[290]

In pursuit of novel antiviral agents that specifically hinder the RNase H activity of HIV-1 RT, our laboratory has looked towards the use of short nucleic acid hairpin aptamers as suppressors of HIV function. The Damha group has had a long vested interest in the discovery of innovative *antisense* therapeutics, which trigger the RNase H-mediated degradation of viral, oncogenic, or mutant forms of mRNA.[377] Since the RNase H activity of HIV-1 RT is absolutely required for completion of the retroviral cycle, hairpin aptamers were selected to explicitly *block* this enzymatic activity from degrading the viral RNA in the DNA/RNA heteroduplex, hence causing hybrid arrest at that stage of viral replication (**Figure 4.3**). A library of structurally and conformationally diverse hairpin molecules based on the highly stabilizing 5'-UUCG-3' tetraloop structure were synthesized by *diversity-oriented* solid-phase synthesis.[378] Of the forty-five hairpins studied, four were shown to potently and specifically inhibit the RNase H activity of HIV-1 RT without affecting its polymerase function or other cellular RNase H's (*e.g. E. coli* and human RNase H).[227] The most potent hairpin inhibitors were composed of native RNA hybrid stems and loops, and displayed IC_{50} values in the 7-30 μM range. Such hairpin constructs demonstrated a global A-type helical arrangement, which has

been shown to be the preferred helical geometry for effective RNase H binding.[290,379] The results revealed the first example of oligonucleotide constructs capable of selectively inhibiting the RNase H functionality of HIV-1 RT without affecting its polymerase activity.

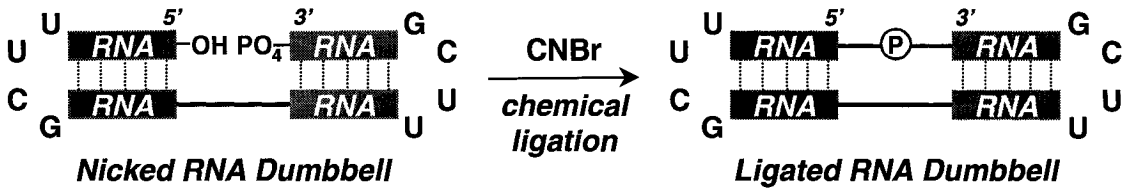


Figure 4.4: Schematic representation of RNA dumbbell ligation using cyanogen bromide (CNBr).

As an extension of this work, we looked at nicked and ligated RNA dumbbells, (*i.e.* double RNA hairpin structures) containing analogous sequences to the most potent RNA hairpins previously studied, and assessed their ability to specifically inhibit the RNase H activity of HIV-1 RT (**Figure 4.4**). The results presented in this chapter are preliminary yet extremely encouraging. In the most potent dumbbell studied, the HIV-1 RT RNase H activity was effectively inhibited without any consequence on its polymerase activity. Moreover, it efficiently suppressed the RNase H activity at a concentration of 3.3 μM.

4.3. CHEMICAL SYNTHESIS OF RNA DUMBBELLS

Recently, it has been demonstrated that 18 base pair heteroduplexes which adopt a predominant A-form helical organization (*e.g.* RNA/RNA or RNA/2',5'-RNA) are capable of binding to HIV-1 RT RNase H and sequester its ability to degrade the RNA strand in an RNA/DNA hybrid.[290] Although, high affinity binding was observed *in vitro*, the bimolecular nature of such complexes makes them difficult to develop into effective therapeutics since it is highly likely that they would not remain in their hybrid state following administration. The RNA/RNA and RNA/2',5'-RNA hybrids were also capable of inhibiting *E. coli* RNase H activity, suggesting that such complexes may

undesirably impede cellular RNase H function. Furthermore, the presence of free termini renders them exceedingly susceptible to degradation by ubiquitous cellular nucleases, predominantly of the 3'-exonuclease type.[380] Similarly, RNA hairpin molecules which adopted global A-type helices were the most potent inhibitors of HIV-1 RT RNase H.[227] These unimolecular complexes displayed high thermal stability[381], nonetheless, the presence of free termini makes them attractive substrates for cellular nucleases, thereby limiting their bioavailability. Several stabilization methods for phosphodiester oligonucleotides have been proposed, such as the incorporation of chemical substituents at the 3'-hydroxyl group[382], the formation of hairpin loop structures at the 3'-end[383,384], or the circularization of the oligonucleotides by joining the 3' and 5'-ends.[247,255] Nucleic acid dumbbells contain termini that are tied up in a circularized structure rendering them resistant to exonucleolytic hydrolysis (**Figure 4.4**). Additionally, the high thermal stability of the complexes imparted by the presence of stabilizing loop structures, should increase the effective concentration of duplexed material upon delivery.

The nicked RNA dumbbells were synthesized on an ABI 381A DNA synthesizer using standard silyl phosphoramidite chemistry and reagents (**Table 4.1**). Terminal 3'-phosphates were incorporated using the chemical phosphorylation reagent as described previously (see section 3.2.2). The UUCG loop sequence was maintained in all dumbbells since it has been demonstrated that this uniquely folded and rigid structural entity[385,386] may be a key recognition motif for effective binding to HIV-1 RT RNase H.[227] Sequences were purified by denaturing PAGE (8.3 M urea), desalted by SEC and their nucleotide composition confirmed by MALDI-TOF-MS.

Chemical ligation of the nicked phosphate hydroxyl junction was conducted using cyanogen bromide (CNBr) as a condensing reagent as described previously (see section 3.2.4). Briefly, nicked RNA dumbbells were dissolved in 0.25 M MES buffer, and heated at high temperature to disrupt any intramolecular secondary structures. The mixtures were cooled slowly to 4°C in order to promote intramolecular association,

Table 4.1: RNA Dumbbell Precursors, T_m Data and Chemical Ligation Yields

Code	Sequence (5'→3')	T_m (°C)	Ligation Yield (%)
4.1	^{HO}GGAC(UUCG)GUCCAAAC(UUCG)GUUU$_P$	44.1, 79.7	0
4.2	^{HO}tGGAC(UUCG)GUCCAAAAAC(UUCG)GUUUt$_P$	43.0, 76.6	77

T_m values are the average of 3 successive runs and are within ±0.5°C (Buffer: 10 mM Na_2HPO_4, 0.1 mM Na_2EDTA, pH 7.0). The ligation yield was determined by densitometric gel quantitation of the circularized product band using the UN-SCAN-IT (Silk Scientific) software program.

whereupon CNBr was added to seal the phosphate/hydroxyl nick-site, thereby forming a circularized double-helical structure (**Figure 4.4**). The extent of ligation in each of the dumbbells was monitored by denaturing PAGE (**Figure 4.5**) and/or reverse-phase HPLC (**Figure 4.6**). As mentioned previously, the nature of the nucleotide residues facing the nicked junction is of utmost importance for realization of high ligation efficiency in dumbbell precursors (section 3.2.4).[312,313] When a 5'-rG/3'-p(rU) was incorporated at the ligation juncture (**4.1**), no circularization was evident, as established by the absence of any new product species on the 16% denaturing gel (**Figure 4.5**). Higher crosslinked gels

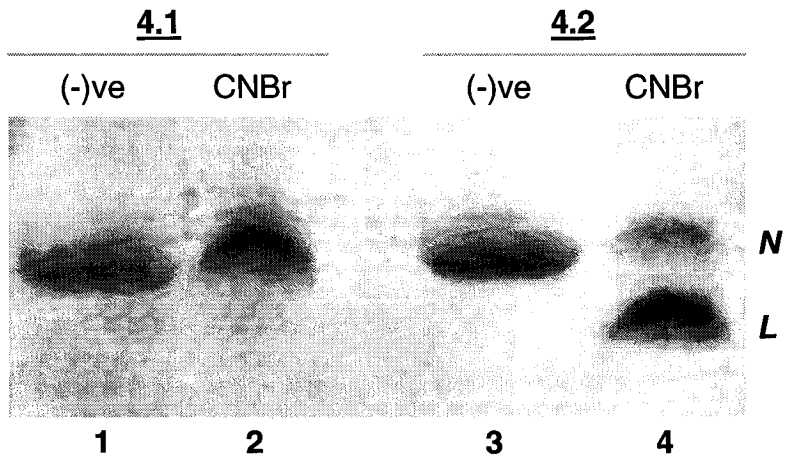


Figure 4.5: 16% PAGE (8.3 M urea) analysis of the CNBr-mediated chemical ligation of **4.1** and **4.2**. Lanes 1 and 3: negative controls consisting of nicked dumbbell only. Lanes 2 and 4: CNBr ligation of nicked dumbbells. *N*= nicked dumbbell precursor. *L*=ligated dumbbell product.

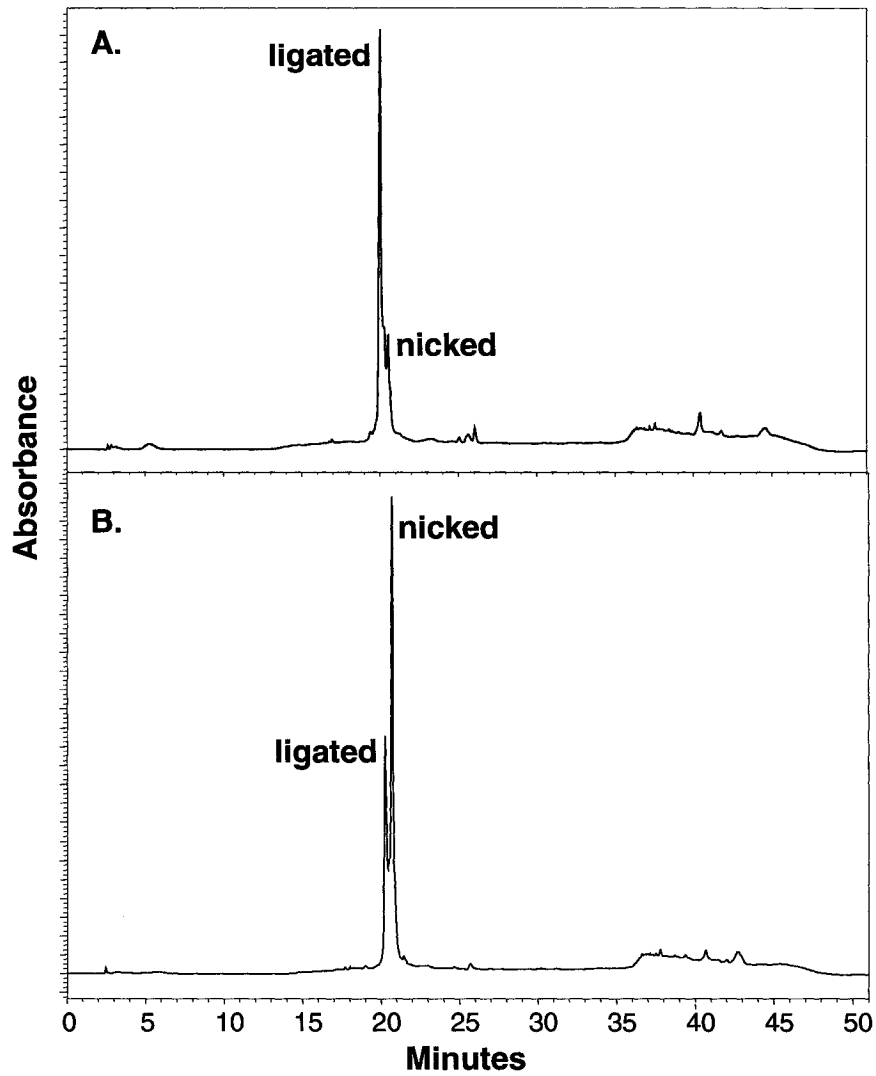


Figure 4.6: Reverse-phase (C_{18}) HPLC analysis of nicked and ligated dumbbell **4.2**. Panel A: chromatogram of crude CNBr ligated **4.2**. Panel B: co-injection of crude CNBr ligated **4.2** and pure nicked **4.2**. HPLC conditions: Buffer A: 0.1 M TEAA, pH 6.5; Buffer B: CH_3CN. Gradient: 100% A over 5 minutes then 0-30% B over 25 min. Column temperature=60°C.

(*i.e.* 20 and 24 %) also confirmed that ligation did not transpire (not shown). This indicates that the phosphate/hydroxyl impingement was not in the appropriate spatial alignment for in-line attack and chemical condensation. Expectedly, this is due to the predominant C3'-*endo* conformation adopted by the ribonucleoside units at the terminal positions. This being the case, the 3'-phosphate would be placed in a pseudoequatorial arrangement, which may sterically hinder its interaction with the neighboring hydroxyl

group. Furthermore, since the 3'-phosphate was adjacent to a reactive 2'-hydroxyl in the ribonucleotide unit, intramolecular 2',3'-cyclophosphate formation may occur upon CNBr activation.[310,323] Potentially, this would lead to a mixture of 2',5' and 3',5'-phosphodiester linkages at the ligation site. However, if the correct local geometry for effective ligation was not achieved, then water would compete with the 5'-hydroxyl for the cyclophosphate, and reversion to the 2' or 3'-phosphate termini would predominate. Given the lack of any new product species, it appeared that the latter prevailed. The terminal ribonucleotide units were then substituted with a more productive 5'-T/3'-pT deoxynucleotide nicked junction as described previously in section 3.2.4 (**4.2**). Indeed, circularization proceeded to afford a 77% yield of cyclized dumbell **4.2** suggesting that the reactive phosphate and hydroxyl units were in the correct stereogeometry for effective ligation (**Figure 4.4**). As demonstrated previously for the DNA dumbbells, the circularized dumbbell displays an accelerated electrophoretic mobility compared to its nicked correlative, likely due to its more compact and globular structure. The new product band was excised from the gel, soaked in water overnight and desalted by SEC. Chromatographic analysis of the ligation mixture by reverse-phase HPLC further demonstrated that circularization was successful (**Figure 4.6**). The closed, circular product had a slightly faster retention time than its nicked counterpart, as a result of its more spherical and compressed structural nature.

Thermal denaturation analysis of the nicked complexes indicated that the intramolecular dumbbell structures **4.1** and **4.2** melted with biphasic profiles (**Table 4.2**, **Figure 4.7**). Under identical buffer conditions (10 mM Na_2HPO_4, 0.1 mM Na_2EDTA, pH 7.0), the open dumbbells exhibited similar T_m transitions as two previously studied independent hairpin structures (**HP-S1** and **HP-S2**)[227], which constituted the nicked dumbbell **4.1** (**Table 4.2**). The large disparity in T_m (>30°C) between the left and right portions of the nicked dumbbells is attributed to one half being comprised of a more thermally stable rG-rC rich stem, whereas the other bisection included an abundant rA-rU hybrid region. The closed, circular nature of ligated dumbbell **4.2** was further confirmed by monitoring its T_m profile (**Figure 4.7**). Indeed, the ligated dumbbell displayed a cooperative, unimolecular order-disorder transition that was significantly higher than the independent transitions

observed for the open dumbbell complex. MALDI-TOF-MS analysis of the pure ligated product exposed a molecular weight consistent with the loss of a water molecule, indicating that the phosphate/hydroxyl junction had been sealed off in the form of a new phosphodiester linkage (see Experimental **Table 7.4**).

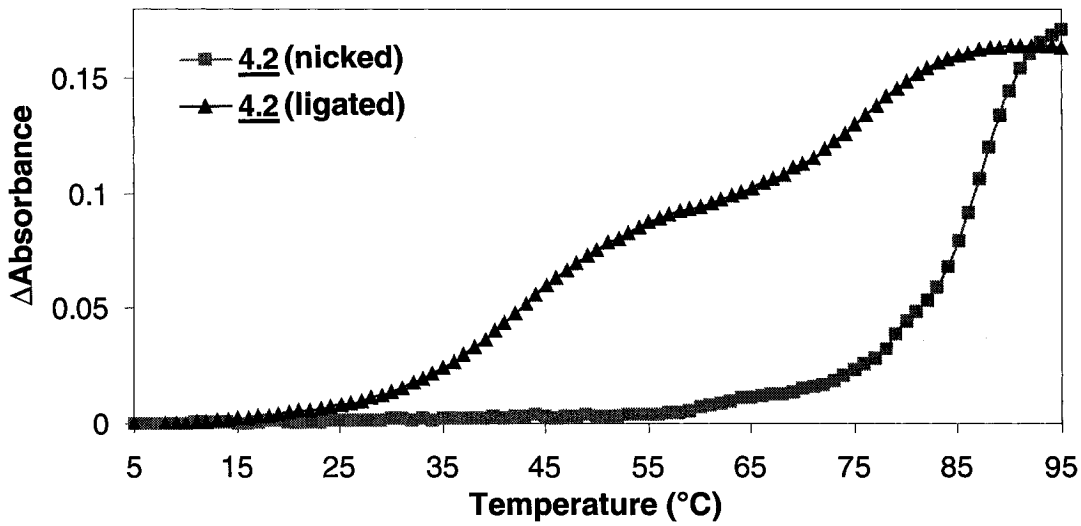


Figure 4.7: Characterization of the ligated RNA dumbbell structure **4.2** by T_m analysis. Buffer: 10 mM $NaHPO_4$, 0.1 mM Na_2EDTA pH 7.0.

4.4. SPECIFIC INHIBITION OF THE RNASE H ACTIVITY OF HIV-1 RT

4.4.1. Comparison of the HIV-1 RT RNase H Inhibitory Activity of RNA Dumbbells *versus* RNA Hairpins

Given the structural and conformational similarities between the nicked (**4.1** and **4.2**) and ligated (**4.2**) forms of the RNA dumbbells and the previously studied RNA hairpins, the dumbbells were tested for their potential inhibitory activity against HIV-1 RT RNase H. The idea of using dumbbell structures as modulators of HIV-1 proliferation is not a new concept. Chimeric RNA/DNA dumbbell oligonucleotides bearing a sense RNA and antisense DNA strand linked by two alkyl loop structures have been investigated for their ability to inhibit HIV replication.[260] Specifically, the constructs were designed to bear

an antisense DNA oligonucleotide, complementary to the HIV-1 *gag* RNA sequence, which was hybridized to a complementary RNA oligonucleotide in the dumbbell structure. Upon delivery into the retrovirus-infected cells, cellular RNase H degraded the RNA portion of the dumbbell, thereby releasing the antisense DNA. The liberated antisense molecule can then go on and hybridize to its complementary target viral RNA, thereby invoking the RNase H mediated degradation of the viral RNA strand. Although highly effective at blocking viral proliferation, the mechanism of action of these chimeric dumbbells was not to inhibit a specific enzymatic function during HIV replication, but rather to target viral gene expression using an antisense mechanism of action. Alternatively, circular dumbbell oligonucleotides have also demonstrated significant biological relevance as aptamers or decoys for hybridizing proteins such as transcription factors.[255,256] Importantly, the circular dumbbell oligonucleotides exhibited relatively high nuclease resistance as well as increased cellular uptake compared to their nicked and linear counterparts.[260,387,388]

In collaboration with Dr. Kyung-Lyum Min of our research group, RNA dumbbells **4.1** and **4.2** were tested for their specific capacity to act as inhibitors of HIV-1 RT RNase H, and their activity compared to the most potent RNA hairpin inhibitors (**HP-S1** and **HP-L**) previously recognized in our laboratory (**Table 4.2**). Initially, a 5'-[^{32}P]-terminally radiolabeled RNA oligonucleotide (18-nt) was annealed to its complementary DNA strand to form a 5'-[^{32}P]-RNA/DNA hybrid. The resultant duplex was treated with HIV reverse transcriptase (p66/p51 heterodimer), which had been pre-incubated with variable concentrations of RNA dumbbell (**Figure 4.8**). The extent of inhibition of the HIV-1 RT RNase H-mediated degradation of the 5'-[^{32}P]-labeled RNA strand in the heteroduplex by the nicked or ligated RNA dumbbell was measured by densitometric analysis as assessed from the decrease of the full-length RNA substrate. The concentration of dumbbell, which was capable of inhibiting 50 % of the RNase H-mediated destruction of the RNA strand, is termed the IC_{50} value. This value was calculated from a plot of the outstanding, undegraded 5'-[^{32}P]-RNA *versus* the concentration of RNA dumbbell oligonucleotide.

Table 4.2: Inhibition of the RNase H Activity of HIV-1 RT by Nicked and Ligated RNA Dumbbell Oligonucleotides and Comparison to Previously Studied Hairpins

Code	Sequence (5'→3')		T_m (°C)	IC_{50} (μM)
HP-S1[a]	^{HO}GGAC(UUCG)GUCC$_{OH}$		71.8	25.8
HP-S2[a]	^{HO}AAAC(UUCG)GUUU$_{OH}$		52.4	-
HP-L[a]	^{HO}GUGGAC(UUCG)GUCCAC$_{OH}$		n.m.	7.8
4.1	^{HO}GGAC(UUCG)GUCCAAAC(UUCG)GUUU$_P$		44.1, 79.7	>60
4.2	^{HO}tGGAC(UUCG)GUCCAAAAAC(UUCG)GUUUt$_P$	N	43.0, 76.6	40.4
		L	87.0	3.3

The IC_{50} value is the amount of dumbbell or hairpin molecule required to inhibit the HIV-1 RT RNase H mediated degradation of a substrate DNA/RNA hybrid by 50% (see Experimental). [a]Hairpin values were obtained from the thesis of Dr. Rami Hannoush (Ref. 227). T_m values for nicked and ligated dumbbells represent the average of 3 successive runs and are within ±0.5°C (Buffer: 10 mM Na_2HPO_4, 0.1 mM Na_2EDTA, pH 7.0). IC_{50} values are the average of 2-3 independent measurements. The error associated with the IC_{50} is represented by a standard deviation of ±1 μM. RNA residues are represented by capital letters whereas small letters indicate DNA residues. Bracketed residues designate the stabilizing hairpin loop base sequence. OH=terminal hydroxyl; P=terminal phosphate; N=nicked dumbbell; L=ligated dumbbell.

Although the loop sequence was identical in both nicked RNA dumbbell structures **4.1** and **4.2**, the degree of inhibition appeared to correlate directly with stem length (4+4 *versus* 5+5 base-pairs). Increasing the length of the stem by one base pair in each bisection more than doubled the potency of the open dumbbell structure **4.2** compared to **4.1** (**Table 4.2**). A similar trend was visible in the RNA hairpin structures previously studied in our laboratory. Specifically, when the stem was composed of six base pairs (**HP-L**) rather than four base pairs (**HP-S1**) a nearly three-fold enhancement in potency was evident (**Table 4.2**).[227] In addition, the most potent hairpin structure (**HP-L**) demonstrated at least five times the inhibitory activity of the nicked dumbbell **4.2**. Interestingly, the ligated form of dumbbell **4.2** (IC_{50}=3.3 μM) was more than ten times more potent than its nicked counterpart (IC_{50}=40.4 μM). Direct comparison of the inhibitory activities of **HP-L** and ligated **4.2** revealed that the circularized RNA dumbbell was at least two-fold more active than the hairpin structure under identical reaction

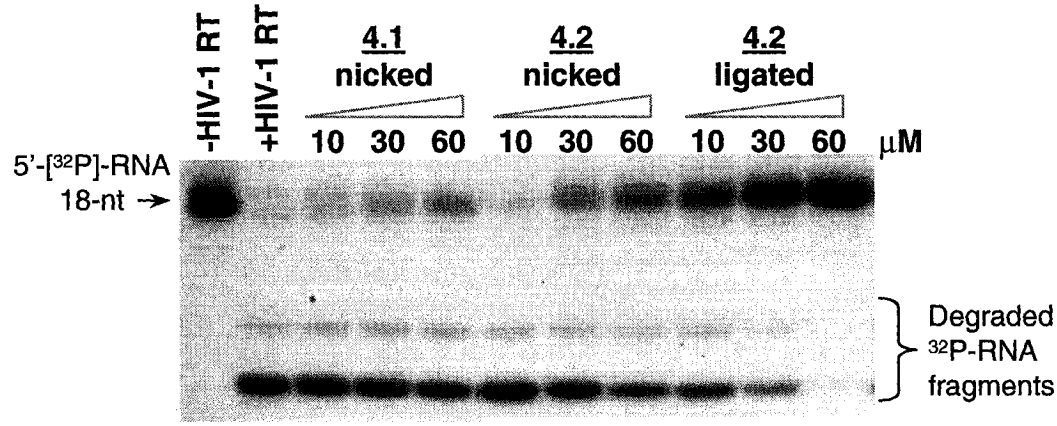


Figure 4.8: Autoradiogram demonstrating the RNase H inhibitory potency of nicked and ligated RNA dumbbells **4.1** and **4.2** in the presence of HIV-1 RT. HIV-1 RT was pre-incubated with 0-60μM of cold nicked or ligated dumbbells. The reactions were initiated by the addition of the 5'-[^{32}P]-RNA/DNA heteroduplex substrate (see Experimental).

conditions (**Figure 4.9**). Whether this effect is dependent on the length of the stem region alone, or on the presence of two, rather than one UUCG loop motif remains to be tested. At low concentration of inhibitors (*i.e.* 5 μM), neither the nicked form of **4.2** or the most potent short RNA hairpin structure (**HP-S1**) displayed any inhibitory activity (**Figure 4.10**). Conversely, both the longer RNA hairpin (**HP-L**) and the ligated RNA dumbbell (**4.2**) effectively inhibited the HIV-1 RT RNase H-mediated degradation of the RNA strand (**Figure 4.10**). Nonetheless, the activity of the RNA dumbbell **4.2** maintained nearly double the potency of hairpin **HP-L** (**Figure 4.10B**). The combined results seem to indicate that stem-length is an important factor in designing potent inhibitors of HIV-1 RT RNase H. The most potent inhibitor; the closed, circular dumbbell **4.2**, enclosed eight base-paired nucleotides in the stem, suggesting that longer RNA/RNA duplexes may be better accommodated in the RNase H domain of HIV-1 RT. As mentioned previously, studies with RNA hairpin structures have indicated that HIV-1 RT distinguishes and recognizes the unusually folded UUCG loop structure as a signal for binding to its substrate.[227] In fact, base mutations within the loop region (UUCG to UACG) of the RNA hairpins completely abolishes hairpin activity.[227] Incorporating a second stabilizing loop motif by creating a double-hairpin structure did not increase the inhibitory potency, rather, biological activity was severely compromised (**Table 4.2**;

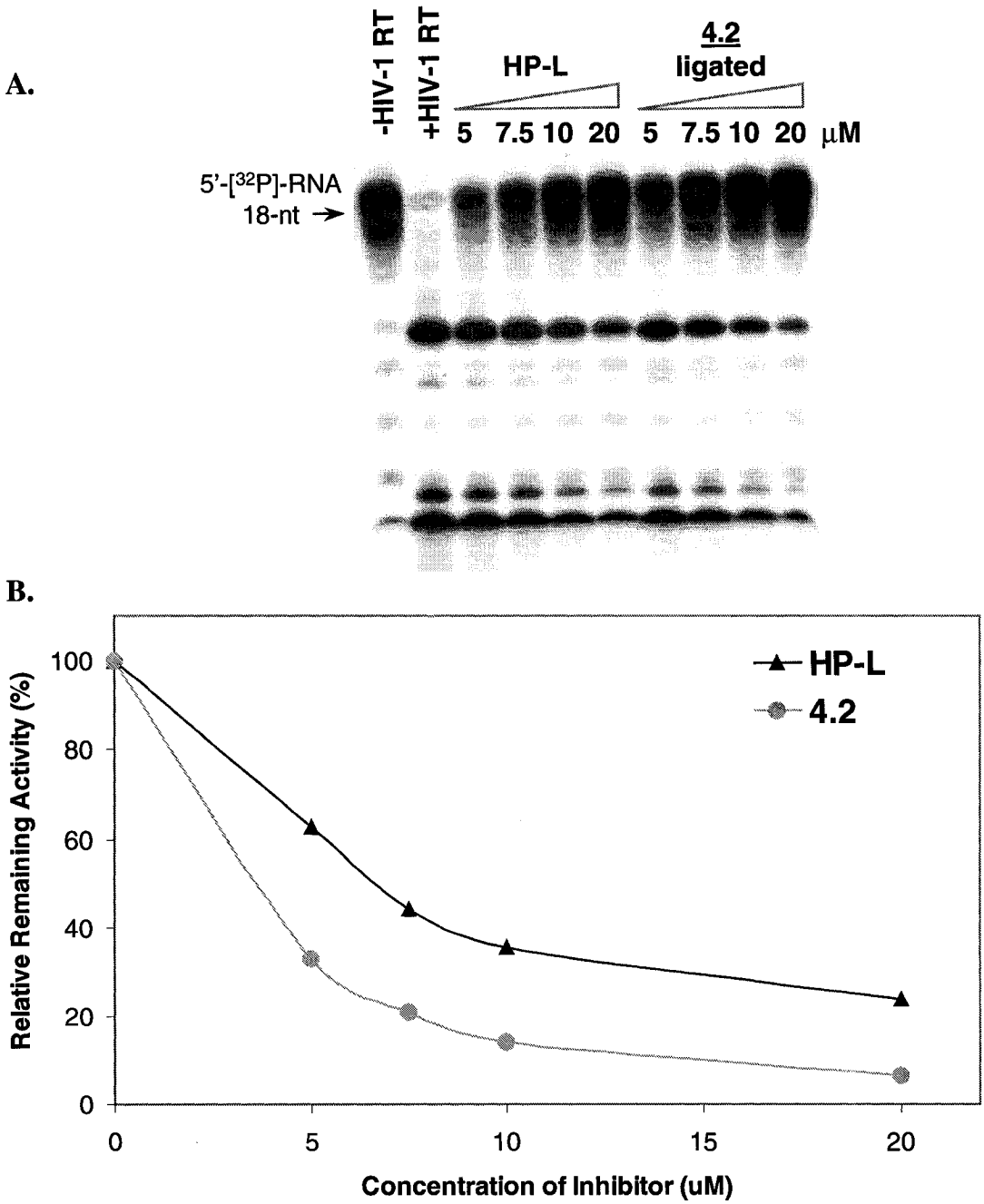


Figure 4.9: Panel A: Autoradiogram comparing the RNase H inhibitory potency of long hairpin RNA (**HP-L**)[227] with ligated RNA dumbbell **4.2** in the presence of HIV-1 RT. HIV-1 RT was pre-incubated with 0-20 μM of cold nicked or ligated dumbbells. Reactions were initiated by the addition of the 5'-[^{32}P]-RNA/DNA heteroduplex substrate (see Experimental). Panel B: Chart demonstrating the inhibitory activity of **HP-L** and RNA dumbbell **4.2**. Inhibitory concentration values were determined by densitometric quantitation of the remaining radiolabeled RNA in the autoradiogram (Panel A)

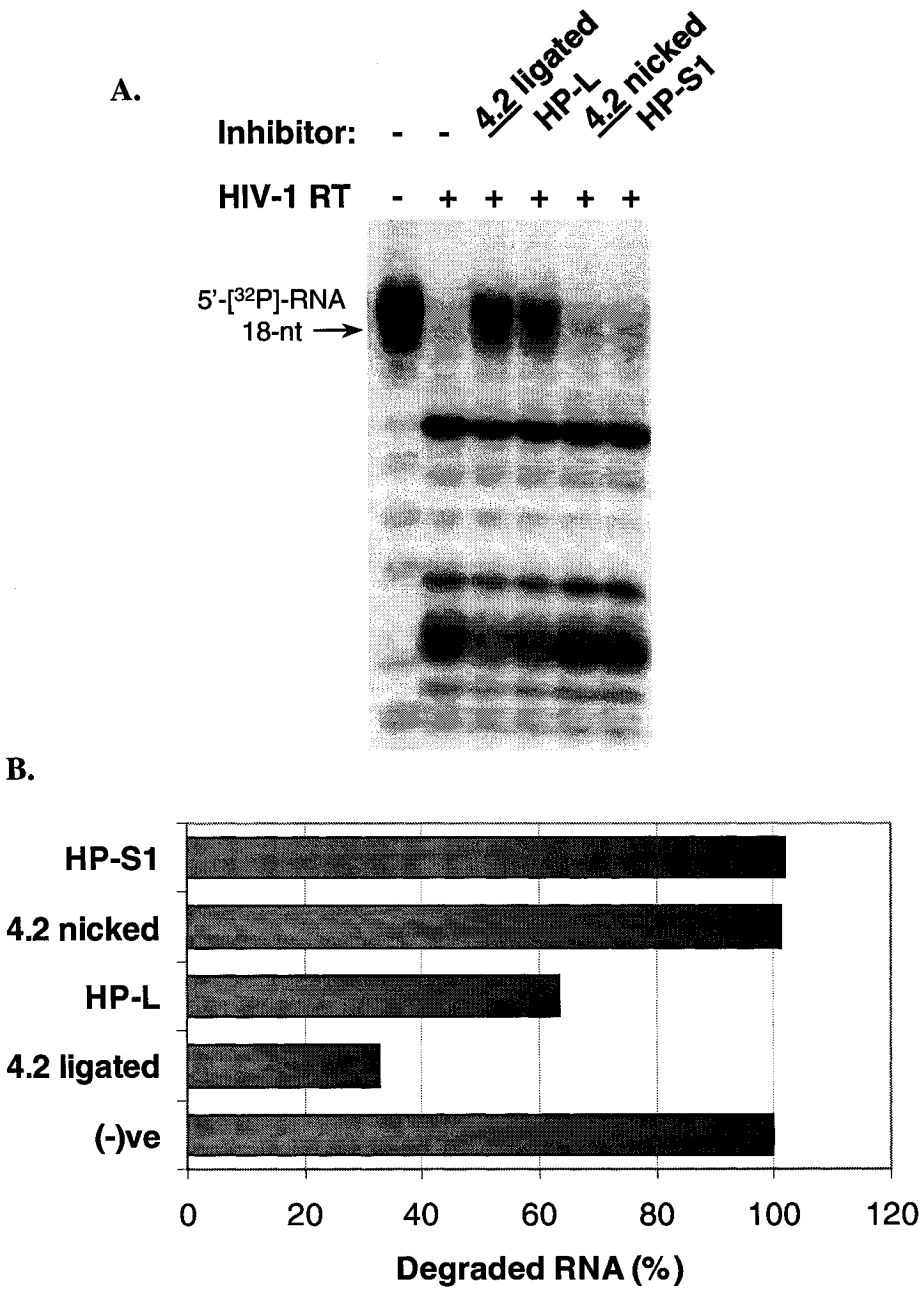


Figure 4.10: Panel A: Autoradiogram comparing the RNase H inhibitory potency of hairpin RNAs, **HP-L** and **HP-S1** (Ref. 227) with nicked and ligated RNA dumbbell **4.2** at low concentration of inhibitor (5 μM). HIV-1 RT was pre-incubated with 5 μM of cold hairpins, nicked dumbbell or ligated dumbbell. Reactions were initiated by the addition of the 5'-[^{32}P]-RNA/DNA heteroduplex substrate (see Experimental). Panel B: Chart demonstrating the inhibitory potency of hairpin and dumbbell RNAs at low concentration of inhibitor (5 μM). The % degraded RNA was determined by densitometric quantitation of the remaining radiolabeled RNA in the autoradiogram (Panel A).

compare **HP-S1** to nicked dumbbells **4.1** and **4.2**). Contrarily, the ligated dumbbell **4.2**, which also contained two UUCG loop motifs, was the most potent of all the molecules studied. This seems to indicate that HIV-RT requires only one loop structure for ample recognition and binding, but a longer stem region is requisite for grasping and positioning the substrate within its binding domain. Regardless, the second loop motif may play an essential role *in vivo*, by stabilizing the structures against the ever-prominent exonucleases present in the biological milieu.

Although the results appear very promising for the development of RNA dumbbells as aptamers against the RNase H functionality of HIV-1 RT, it still remained to be shown that such structures are specific to this ortholog of the enzyme. For example, a few HIV-1 RT RNase H inhibitors such as the RNA/RNA and RNA/2',5'-RNA hybrids, as well as the natural product illimaquinone, were capable of inhibiting HIV-1 RT RNase H as well as *E. coli* RNase H activity, suggesting that such molecules may undesirably obstruct cellular RNase H activity. Although the three-dimensional structure of a eukaryotic RNase H has not yet been described, it is possible that structural similarities exist, as with HIV-1 and *E. coli* RNase H. This similarity may hamper the use of retroviral RNase H as a target to stop retroviral replication without affecting cellular functions, unless the RNase H inhibitor is specific only to the HIV-1 homolog. In an effort to verify the specificity of HIV-1 RT RNase H inhibition and to substantiate the potential use of such dumbbell structures as chemotherapeutic agents for HIV-1 infection, we tested their ability to inhibit both bacterial and mammalian homologs of RNase H. The identical 5'-[^{32}P]-RNA/DNA heteroduplex used in the inhibition profiles above was utilized in this assay. Duplexes were incubated with *E. coli* RNase H or human RNase H (type II) in either the absence or presence of cold hairpin (**HP-L**) or ligated RNA dumbbell (**4.2**) inhibitors (**Figure 4.11**). RNase H inhibition was monitored by comparing the amount of intact RNA present in the reaction, to that lacking any inhibitor. Formerly, the RNA hairpin structure **HP-L** demonstrated no specificity towards either homolog of RNase H.[227] This was reproducibly confirmed in the assay shown (**Figure 4.11**; lanes 3 and 6). Similarly, RNA dumbbell **4.2** did not effect either the bacterial or human RNase H-

mediated degradation of the RNA template strand, indicating a remarkably specific effect toward the retroviral RNase H domain.

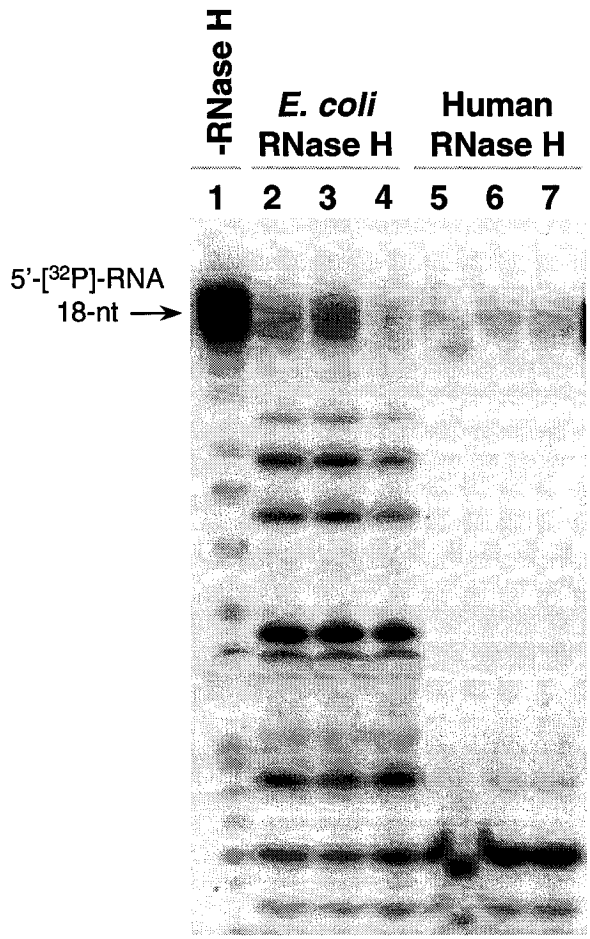


Figure 4.11: Inhibition of RNase H activity by ligated RNA dumbbell **4.2** is specific to HIV-1 RT. Dumbbell **4.2** was tested alongside long hairpin **HP-L** (Ref. 227) for its inhibitory potential of both bacterial RNase H (*E. coli*) and Human RNase H (type II). Homologs of RNase H were pre-incubated with either 60 μM of cold hairpin (**HP-L**) or ligated dumbbell (**4.2**). The reactions were initiated by the addition of the 5'-[^{32}P]-RNA/DNA heteroduplex substrate (see Experimental) and the amount of degraded radiolabeled substrate RNA quantified using the UN-SCAN-IT software program. Lane 1: negative control(no enzyme); Lane 2: duplex + *E.coli* RNase H (no inhibitor); Lane 3: duplex + *E. coli* RNase H + **HP-L**; Lane 4: duplex + *E. coli*. RNase H + **4.2**; Lane 5: duplex + human RNase H (no inhibitor); Lane 6: duplex + human RNase H + **HP-L**; Lane 7: duplex + human RNase H + **4.2**.

4.4.2. RNA Dumbbells Do Not Inhibit the DNA-Dependent and RNA-Dependent DNA Polymerase Activities of HIV-1 RT

Specific inhibition of reverse transcription is complicated by the fact that RT has two enzymatic activities on the same substrate; the polymerase domain that copies the viral RNA template, and the RNase H domain that subsequently degrades it. Both enzymatic active sites appear to be spatially and temporally coordinated, with 18 to 19 nucleotide units separating the two activities.[389,390] This being the case, a variety of HIV-1 RT RNase H inhibitors have been disclosed, however, most of these compounds are not specific to this enzymatic activity and inhibit the polymerase functionality as well. Keeping this in mind, the RNA dumbbells were assayed for their ability (or inability) to inhibit DNA synthesis catalyzed by HIV-1 RT. In order to exclude the possibility that the RNA dumbbells were acting as polymerase inhibitors, which indirectly affected the RNase H function of RT, the compounds were tested directly against the DNA polymerase activity of the enzyme. The effects on both the RNA-dependent and DNA-dependent polymerase reactions were tested using either a 5'-[^{32}P]-DNA primer:RNA template or a 5'-[^{32}P]-DNA primer:DNA template complex respectively. The 30-nucleotide template DNA and RNA corresponded to the primer binding site (PBS) sequence of the HIV-1 genome, while the terminally radiolabeled DNA primer was complementary to the 3'-end of the PBS RNA or DNA. Complementary DNA strand synthesis was initiated by the addition of the annealed primer:template complex and various nucleoside triphosphates to HIV-1 RT, which had been pre-incubated with cold RNA dumbbell **4.1** and **4.2** (**Figure 4.12**). In the absence of inhibitor, a new, full-length DNA oligonucleotide product was formed from both the RNA primer (RDDP activity; lane 3) and DNA primer (DDDP activity; lane 2). Owing to its higher mass to charge ratio, the newly synthesized DNA complement (30-nucleotides) exhibits a retarded electrophoretic mobility on a denaturing gel compared to its corresponding 18-nucleotide primer (**Figure 4.12**). In all cases, the band corresponding to the full-length complementary DNA product synthesized by the DDDP activity was highly diffused.

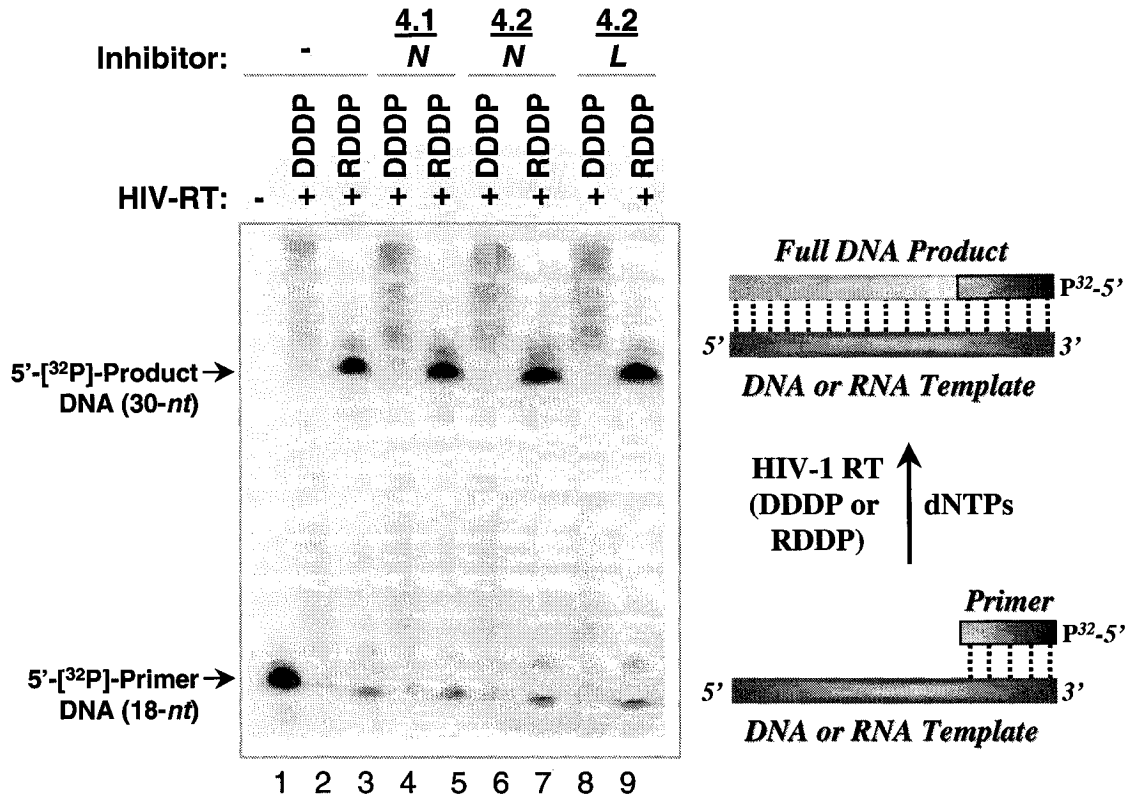


Figure 4.12: Inhibition of DNA synthesis catalyzed by HIV-1 RT DNA-dependent (DDDP) or RNA-dependent (RDDP) DNA polymerase activity by nicked and ligated RNA dumbbells **4.1** and **4.2**. Cold RNA dumbbells (80 μM) were pre-incubated with HIV-1 RT at room temperature for 20 min prior to the addition of either a 5'-[^{32}P]-Primer:DNA template or 5'-[^{32}P]-Primer:RNA template complex and dNTPs. HIV-1 RT catalyzes the synthesis of a complementary DNA strand (5'→3'direction) in the presence of deoxynucleoside triphosphates (dNTPs). Polymerization reactions were conducted for 15 minutes both in the absence (-) and presence (+) of cold RNA dumbbells.

Indeed, the results obtained indicated that both the DNA-dependent and RNA-dependent DNA polymerase functionalities of HIV-1 RT were not compromised in the presence of either the nicked or ligated forms of the RNA dumbbells. Specifically, the intensity of the product DNA bands produced by the DDDP and RDDP activities in the presence of inhibitors were similar to those reactions that lacked any inhibitor (**Figure 4.12**). These results indicate that the RNA dumbbell molecules do indeed act as aptamers toward HIV-1 RT and most importantly, they inhibit the RNase H functionality without any consequence on the polymerase activity. Although the polymerase activity remains unaffected by the dumbbell inhibitors, it still remained unclear whether the aptamers were

binding directly to the RNase H domain of the p66/p51 subunit. This will be addressed in the next section.

4.4.3. UV-Crosslinking of an RNA Dumbbell to HIV-1 RT RNase H

In order to confirm that the RNA dumbbell aptamers were binding specifically to the RNase H domain of HIV-RT, we conducted a UV-crosslinking assay between the most potent dumbbell, ligated **4.2** and either the HIV-1 RT heterodimer (p66/p51), containing both the DNA polymerase and RNase H domains, or the homodimer (p51/p51) consisting of only a functional DNA polymerase domain. As mentioned previously, the p66 monomeric subunit of HIV-1 RT is preoteolytically processed to form both a p51 and p15 subunit, resulting in an RNase H deficient product.[391] In the virion particle, p66 is always found in stable association with the p51 subunit, and it is this resultant heterodimeric entity that displays the full functionality of the HIV-1 reverse transcriptase.[392,393] By taking advantage of the natural photoreactivity of the RNA bases at 254 nm it may be possible to form a stable, localized complex (*i.e.* crosslink) between the RNA dumbbell aptamers and the RNase H domain located in the C-terminal portion of the p66 subunit. Generally, when a covalent crosslink is formed between an enzyme and its oligonucleotide substrate, the complex exhibits an altered and retarded electrophoretic mobility compared to its unbound state. Typically, the oligonucleotide substrate is radioactively labeled so that a complex can be easily visualized by autoradiography. However, in the case of the RNA dumbbells, both the 5' and 3'-termini of the molecule are engaged in a circularized structure, so introduction of a terminal radiolabel was unfeasible. Alternatively, complex formation could also be discerned by monitoring a change in the electrophoretic mobility of the protein subunits themselves on a denaturing sodium dodecyl sulfate (SDS) gel followed by staining of the protein complex.

Ligated RNA dumbbell **4.2** was incubated with either the fully functional p66/p51 HIV-1 RT dimer or the RNase H deficient p51/p51 dimer for a 30-minute period at physiological

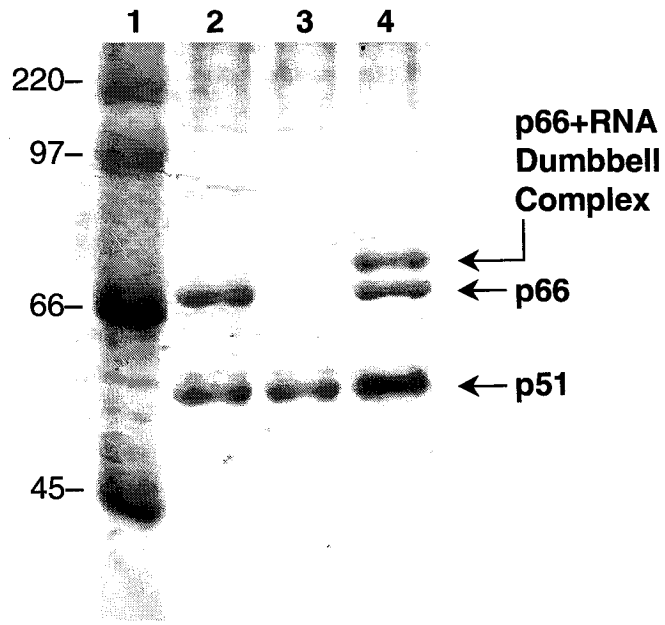


Figure 4.13: UV crosslinking analysis of HIV-1 RT with ligated RNA dumbbell **4.2**. Heterodimeric (p66/p51) and homodimeric (p51/p51) RT were incubated with unlabeled RNA dumbbell (50 pmol) and irradiated with UV light (λ=254 nm) for 15 min on ice. Protein subunits and complexes were partitioned on a 12% SDS-PAGE gel and stained with Coomassie® Brilliant Blue G-perchloric acid solution. Lane 1: molecular weight markers (in KDa); Lane 2: HIV-1 RT p66/p51 heterodimer (no dumbbell inhibitor); Lane 3: HIV-1 RT p51/p51 homodimer (only polymerase domain-no RNase H domain) irradiated with UV in the presence of RNA dumbbell **4.2**; Lane 4: HIV-1 RT p66/p51 heterodimer (both polymerase domain and RNase H domain) irradiated with UV in the presence of RNA dumbbell **4.2**.

temperature (37°C). The oligonucleotide-enzyme mixtures were placed on ice, so as to keep the complex stabilized, and irradiated with UV light (λ=254 nm) for 15 minutes. Complexes were then partitioned on a 12% SDS-PAGE, and stained with Coomassie® blue solution (**Figure 4.13**). The results clearly demonstrate that the circular RNA dumbbell does not form a covalent complex with the p51/p51 homodimer, which lacks the RNase H binding domain (**Figure 4.13**; lane 3). Since the p51/p51 homodimer possesses functional DNA polymerase activity, it is expected that if the RNA dumbbell recognizes this domain, then a stable adduct would form. Conversely, aggregate formation with the p66/p51 heterodimeric species was evident, as evidenced by the presence of a slower migrating product complex on the SDS-PAGE (**Figure 4.13**; lane 4).

Furthermore, the covalent complex formed between the p66 subunit and the ligated RNA dumbbell displayed a molecular weight consistent with the expected (*ca.* 75 KDa). These findings clearly ascertain that the RNA dumbbell aptamer **4.2** does not bind the DNA polymerase region of HIV-1 RT, and instead, is highly specific toward the RNase H domain of the enzyme. This is further corroborated by the results in the previous section, which clearly revealed that RNA dumbbell **4.2** does not have any effect on the HIV-1 RT-mediated synthesis of DNA by the DDDP or RDDP activities. The nature of binding of the RNA dumbbell, whether in the active site of the enzyme (competitive inhibition) or at a secondary site (non-competitive inhibitor) thereby inducing an allosteric change in the enzyme active site, has not been established thus far.

4.5. CONCLUSIONS

The results presented herein disclose a novel example of an oligonucleotide-based inhibitor of HIV-1 RT, an RNA dumbbell, which is capable of selectively hampering the RNase H-mediated activity of the retroviral enzyme. Specifically, such constructs were shown to effectively inhibit the RNase H functionality of RT without any consequence on its polymerase activity. The most potent construct, a ligated RNA dumbbell consisting of a 10 base-pair stem and two flanking UUCG loop motifs, demonstrated extremely potent inhibition of RNase H with an IC_{50} in the 3 μM range.

A high yield of circularized RNA dumbbell was achieved when an extremely non-productive 5'-rG/3'p(rU) junction was replaced with a more proficient 5'-T/3'-pT ligation junction, demonstrating once again that the local stereogeometry at the nick site is vital for controlling the effectiveness of chemical condensation. In addition, incorporation of two adjoining deoxynucleoside residues ensures natural 3',5'-phosphodiester linkage purity at the ligation site by preventing the formation of a 2',3'-cyclophosphate which could form if ribonucleoside residues were used instead. The open dumbbells exhibited biphasic thermal melting profiles, consistent with the presence of nicked double hairpin structures with diverse sequence composition. Alternatively, thermal dissociation of the

ligated dumbbell **4.2** displayed a cooperative, unimolecular transition indicating that cyclization had indeed been achieved.

Nicked and ligated RNA dumbbells **4.1** and **4.2** were tested for their ability to inhibit the RNase H mediated degradation of the RNA strand in an RNA/DNA heteroduplex. Both the open and closed forms of the dumbbells hindered the RNase H activity of HIV-1 RT, however, maximal inhibition was observed in those structures that contained longer stem regions. The results were consistent with previous data demonstrating that hairpin RNA, containing the identical UUCG loop motif, but with longer stem portions, were more effective at inhibiting enzymatic function. In addition, the inclusion of a second UUCG loop in the open dumbbell structures appeared to have an injurious effect on inhibitory activity, suggesting that only one loop is required for ample recognition by the enzyme and that at longer duplex region is more suitably accommodated in the binding domain. Although, nuclease stability assays were not conducted on the dumbbell structures at the present time, previous evidence supports the notion that such entities would be highly stabilized against cellular nucleases making them good candidates for *in vivo* studies. Furthermore, HIV-1 RT DNA polymerase activity was not compromised by the addition of either nicked or ligated RNA dumbbells, demonstrating a specific affinity for the RNase H domain of the enzyme. This was further supported by a UV-crosslinking assay, which clearly established that the most potent RNA dumbbell aptamer (ligated **4.2**) was specifically interacting with the RNase H region of the RT enzyme. Importantly, ligated **4.2** did not inflict any effect on mammalian RNase H activity (human RNase H type II), signifying that such compounds would not interfere with cellular RNase H function should they be developed into chemotherapeutic agents.

The preliminary results clearly imply that RNA dumbbells can potentially be used as selective mediators of HIV-1 expression by blocking the RNase H activity of the retroviral RT enzyme without obstructing its polymerase activity. Currently, the nuclease resistance of the RNA dumbbells in biological media is being explored. In addition, cell culture studies are being pursued, thereby affording a more conclusive indication as to the inhibitory activity of these constructs in HIV infected cells.

CHAPTER 5: INHIBITION AND MODULATION OF PRE-mRNA SPLICING USING SYNTHETIC BRANCHED NUCLEIC ACIDS (bNAs)

5.1. INTRODUCTION

Synthetic oligonucleotides have demonstrated extensive and divergent uses as probes in both molecular biology and for the study of numerous biochemical processes. Linear constructs have proven expedient for the study of nucleic acid structure and function as well as to gain insight into the multitude of protein-nucleic acids interactions, which arise in a variety of organisms. Many of these dynamic interactions occur during pre-mRNA splicing, and as such, synthetic nucleic acid constructs have been central to investigating key events during splicing assembly and reactions. Distinctively, chiral phosphorothioate modifications within splicing substrates have been indispensable for uncovering the stereochemical course of each transesterification step of the splicing reaction.[95,96] Other nuclease resistant oligonucleotides, such as 2'-O-allyl and 2'-O-methyl modifications, have been extensively effective as complementary antisense probes for the isolation of ribonucleoprotein (RNP) complexes, specifically snRNAs.[394,395] Due to the ease with which oligonucleotides can now be prepared, the antisense approach can rival over immunological approaches (*i.e.* antibody selection) for the characterization RNP structure and function.[394] "Caged" RNA molecules have also been successfully used in the study of spliceosomal assembly events.[396] Moreover, synthetic branched oligonucleotides, comprising of vicinal 2',5' and 3',5'-phosphodiester linkages can now be made relatively simply and inexpensively, thereby allowing for a systematic evaluation of such molecules in terms of branch-point recognition by pre-mRNA splicing factors in their native splicing system. Section 5.2 will highlight some of the key results obtained in our laboratory concerning the utility of branched RNA oligonucleotides for studying branch-point interactions in constructs similar to native RNA lariat structures, as well as methods for stabilizing such molecules against cellular nucleases and RNA lariat debranching activity. Oligonucleotide constructs have also been successfully utilized as explicit regulators of alternatively spliced gene expression. Briefly, alternative splicing is evolution's method of simplifying the number of genes required by a specific organism, such that a particular

pre-mRNA, containing multiple exons and introns can be spliced in different fashions, leading to the production of diverse mRNAs, and hence diverse protein products, emerging from a single gene.[124,125] Specifically, antisense oligonucleotides[397-399], complementary to specific segments of a pre-mRNA transcript, have proven particularly useful for the correction of aberrant splicing in genetic disorders such as β-thalessemia[400,401], cystic fibrosis[402] and Duchenne muscular dystrophy (DMD)[403] both *in vitro* and *in vivo*, thereby restoring normal protein function. Additionally, antisense oligonucleotides have shown remarkable success at redirecting alternative splicing of antagonistic genes such as Bcl-x_L and Bcl-x_S, which are directly related to controlling programmed cell death or apoptosis in oncogenic (*i.e.* cancerous) cells.[140,141] The effects of using linear and branched oligonucleotide constructs containing both antisense and protein-binding sequences as a novel means for upregulating pro-apoptotic Bcl-x_S expression *in vitro* will be described in section 5.4.

5.2 Inhibition of In Vitro Pre-mRNA Splicing by Branched Oligonucleotides

5.2.1. Branchsite Recognition during pre-mRNA Splicing

Nuclear pre-mRNA splicing is a highly dynamic process involving an intricate plethora of RNA-RNA, RNA-protein and protein-protein interactions for correct sequence alignment and sequential transesterification reactions to occur.[49,54,404,405] The snRNAs, especially U2 and U6 are intimately associated with the reactive regions of the intronic segment during the chemical steps of splicing, supporting the hypothesis that the splicing reactions are RNA catalyzed.[80,81,83,84,406] Furthermore, construction of the catalytically active *spliceosome*[50] complex is heavily dependent on the activity of numerous protein factors, of which more than 70-100 have been identified to date.[52,88] Many of these protein factors are homologous to both the lower eukaryotic yeast, and the mammalian splicing systems. Likewise, the majority of the *trans*-acting factors involved in the splicing pathway are also highly conserved and present thoughout the eukaryotic kingdom. While the signals for RNA splicing are more divergent in humans than in lower

eukaryotes, the chemistry of the splicing reaction remains evolutionarily conserved. Particularly, in yeast the branchpoint sequence (BPS); UACUAAC (A=branchpoint adenosine) is highly conserved, whereas in mammalian systems, less sequence stringency is observed with YURAC (Y=pyrimidines; R=purine) as the BPS, however the yeast sequence remains optimal in both systems.[56,57,407] Moreover, the density of introns in their genomes mirrors the complexity of the organism. Specifically, yeast have only a few hundred introns and are generally limited to one intron per gene. Instead, mammalian genes contain multiple introns thereby yielding tens (or hundreds) of thousands of introns coded within their genomes. As such, the snRNA sequences contain all the necessary information for interaction with their substrate RNA, each other, and the auxiliary proteins that direct them to their appropriate positions in the activated splicing complex. Even though substantial research efforts have been directed towards complete elucidation of the splicing process and the synergistic relationship between the multitudes of factors involved, the overall process remains yet to be defined precisely.

Of specific interest to our group is the web of complex interactions that hold the highly conserved branchpoint region together for splicing to occur. It is recognized that the U2 and U6-snRNAs restrain the 5'-splice site and branch region in an accurate spatial arrangement for the first step of splicing to occur, whereby the branched lariat-3'-exon intermediate is formed and the 5'-exon is released for ligation to the 3'-exon during the second transesterification reaction. In both the yeast and mammalian systems, the pre-mRNA BPS is recognized several times during the splicing process. Firstly, the BPS is recognized by a single-stranded RNA-binding protein referred to as the branchpoint binding protein (BBP) in yeast or SF1 (*i.e.* mBBP) in mammals.[408,409] BBP appears to be involved only in early recognition since it is present only during initial spliceosomal assembly events.[410] Subsequently, the U2-snRNP binds to the BPS in part by base-pairing interactions between the U2-snRNA and the nucleotides flanking the branchpoint adenosine unit.[66-68] Rather than being completely base-paired to the U2-snRNA, the branchpoint adenosine residue is speculated to be unpaired and "bulged" out of the duplex region, thereby making it available for ensuing interactions that will position it for nucleophilic attack at the 5'-splice site (see **Figure 1.6**).[69] X-ray crystallographic

analysis of a model BPS-U2-snRNA complex has indicated that such a "bulging" conformation is indeed plausible.[411] In addition, as the spliceosome repositions itself to reach the correct conformation for catalysis, other protein factors have been shown to interact with the BPS *via* site-specific crosslinking interactions with the pre-branched adenosine.[412,413] Allegedly, one of these factors, p14, may play a role in positioning the adenosine moiety for attack at the 5'-splice site junction.[108]

Although at the present time a wealth of information regarding the interactions between various pre-spliceosomal and spliceosomal factors and the BPS adenosine moiety prior to branch (*i.e.* lariat) formation is available, little is known about the events surrounding branchpoint recognition after step one is complete (*i.e.* lariat-3'-exon recognition). Obviously, the chemical involvement of the branch site adenosine is only during step one, and not in the second step (*i.e.* exon ligation); nevertheless, the identity of the adenine base is critical for both steps.[414] Particularly, an adenine to guanine base change imposes a strong block to step two, as well as a considerable decrease to step one[69,103], suggesting that the nature and topology of the branch is important for interactions specific to the second step activation mechanism.[59]

Given that synthetic linear oligonucleotide constructs have proven expedient for the study of pre-mRNA splicing, it seemed wise and worthwhile to investigate branchpoint recognition events by *trans*-acting spliceosomal factors using synthetic branched nucleic acids (bNAs) in their native splicing system. Until Robert Hudson, a previous member of our research group, had initiated an investigation on bNA recognition by yeast nuclear extract during *in vitro* pre-mRNA splicing, bNAs had only previously been used as haptens for the characterization of antibranch antibodies[229] and preliminary efforts had emerged on the use of bNAs to study lariat RNA catabolism by the RNA debranching enzyme, a specific 2'-phosphodiesterase.[115,216] His rudimentary results indicated that fully formed homopolymeric bNAs (Y-DNAs) were capable of inhibiting pre-mRNA splicing in yeast, albeit at high concentrations, suggesting that bNAs were selectively sequestering a branch-recognition factor possibly involved in branch selection after step

one (**Figure 5.1**).[208] Ultimately, the goal of the project was to attempt affinity purification of the spliceosomal (or non-spliceosomal) element(s) that recognizes and binds the bNAs during splicing. Consequently, another former graduate student from our group, Ravi Braich, began to examine the merit of the ribose sugar, bNA sequence and branchpoint nucleotide on bNA recognition by yeast extract during pre-mRNA splicing.[212] Although promising inhibitory profiles were obtained with bRNA molecules, the natural constituents of the bNAs (*i.e.* ribose sugar), especially at the 3'-termini of the molecule, were promoting ample degradation of the substrates by ubiquitous cellular nucleases. In effect, a thorough study was not possible given this limitation.

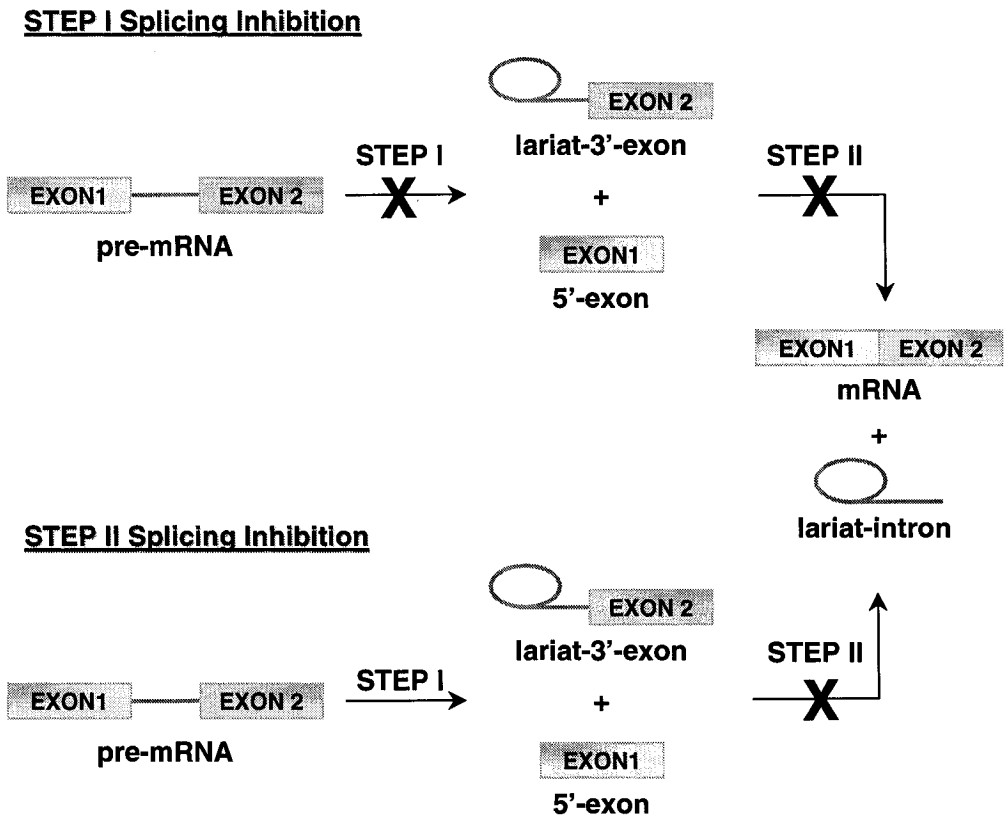


Figure 5.1: Stages of intervention of RNA splicing inhibitors. An inhibitor can either act by blocking Step I (*i.e* branched lariat-3'-exon intermediate formation and 5'-exon release) or Step II (*i.e.* mRNA ligation and lariat-intron release) of the splicing reaction. The specific interruption step can be monitored by denaturing PAGE. Step I inhibition results in a decrease in the amount of both the lariat-3'-exon (Step I product) and lariat-intron (Step II product). Step II inhibition gives rise to an accumulation in lariat-3'-exon (Step I product) and decrease in lariat-intron (Step II product).

In the following sections, efforts to stabilize the termini of the bNAs and study if terminal modifications were still conducive to splicing inhibition, will be described. Studies were first conducted in the yeast system (*S. cerevisiae*; section 5.2.2). Given the high degree of homology between the yeast and mammalian splicing factors, studies were later translated to the more biologically complex mammalian splicing system (section 5.2.3). Additional efforts to stabilize the bNAs against omnipresent endonuclease and debranching activity in the HeLa extract will also be discussed.

5.2.2. Inhibition of *In Vitro* pre-mRNA Splicing in Yeast (*S. cerevisiae*) by Branched RNAs

Since the advent of splicing in 1977[36,37], numerous research labs have attempted to emulate the delicate *in vivo* post-transcriptional events in the test tube (*i.e. in vitro*) using nuclear extracts.[415-417] *In vitro* splicing systems using both yeast (*e.g. S. cerevisiae*) and mammalian (HeLa) nuclear components are now the generality in labs investigating spliceosome structure and function, RNA recombination events and *trans*-splicing as a potential tool for gene therapy.[47,48] Appropriately, the two distinctive transesterification reaction steps, which arise during splicing of an RNA transcript, can be cleanly and easily separated by virtue of gel electrophoresis. Generally, a similar pattern of the splicing intermediates and products emerges on the gel regardless of the RNA splicing substrate utilized. When a reaction is allowed to proceed for a short amount of time, typically 20-30 minutes, the initial pre-mRNA, the step one intermediates (*i.e.* lariat-3'-exon and 5'-exon) and the step two products (*i.e.* lariat intron and mRNA) can be individually identified according to their mobility on the gel (**Figure 5.2**). The abnormal mobility of the lariat-containing species (*i.e* compared to a linear correlative of identical size and base composition) is demonstrative of its unique topological structure consisting of a circular RNA with a tail connected by vicinal 2',5' and 3',5'-phosphodiester linkages at the branched adenosine core.[109,415] As demonstrated in **Figure 5.2**, exonic sequences in the pre-mRNAs under investigation will be represented by boxes, whereas a line will signify a single intronic segment.

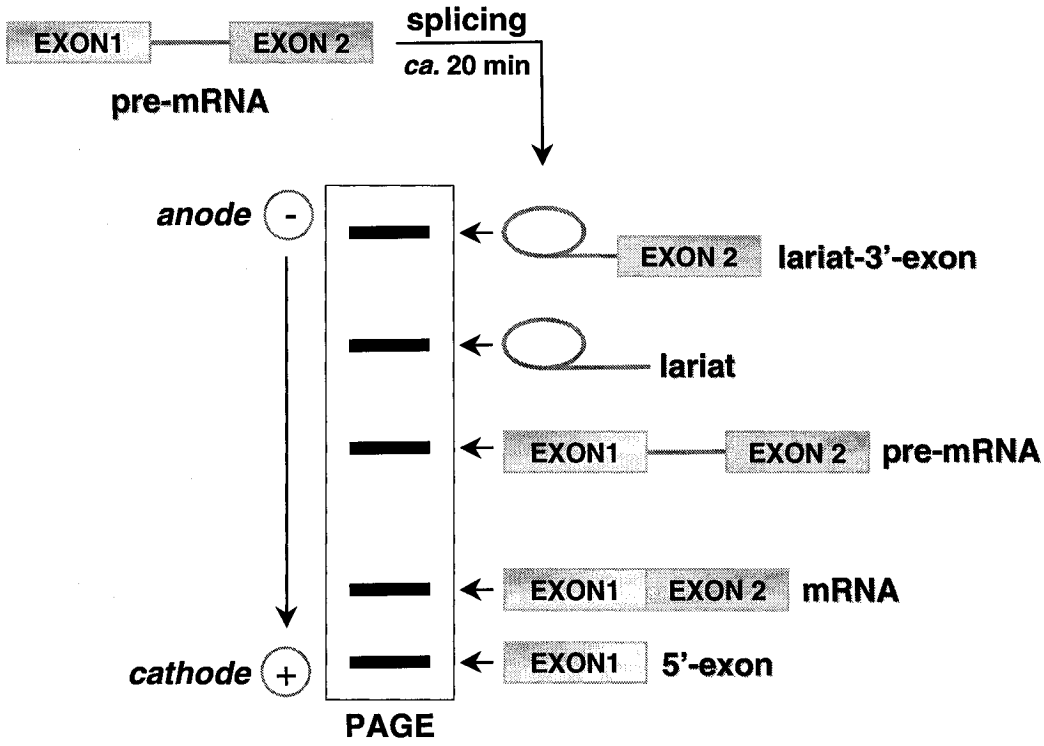


Figure 5.2: General migratory pattern of the pre-mRNA splicing precursor, intermediates and products on a denaturing polyacrylamide gel (PAGE). The directionality of the current is indicated by the arrow (*i.e.* anode→cathode). The slowest moving band (top; most retained) is typically the lariat-3'-exon splicing intermediate whereas the fastest moving (bottom; least retained) is the 5'-exon. Exonic (*i.e.* coding) sequences are represented by boxes, while the intronic (*i.e.* non-coding) sequences are represented by lines.

Yeast nuclear splicing extract was prepared in the laboratory of Dr. James D. Friesen at the University of Toronto (Banting and Best Department of Medical Research) with the cooperation of Dorian Anglin. Whole cell extracts were prepared according to a modified protocol of Schultz and co-workers (see Experimental; section 7.12.1.1).[418] A protease deficient strain of yeast (W3031A) was grown and collected during logarithmic growth. The cells were fractured by freezing with liquid nitrogen and manually homogenized to a fine powder with the aid of a ceramic mortar and pestle, then redissolved in extraction buffer (section 7.12.1.1). The highly soluble fraction was isolated by ultra-centrifugation, dialyzed and supplemented with fresh protease inhibitors. The splicing extract, containing all the components necessary for *in vitro* splicing, was aliquoted into smaller portions and stored at -80°C until needed.

Table 5.1: List of linear, V- and Y-shaped oligonucleotides utilized as inhibitors of *in vitro* pre-mRNA splicing in *S. cerevisiae*

Code	Sequence (5'→3')	Topology	IC_{50} (μM)
Linear Oligonucleotides			
5.1[a]	AUG GAU UCU GAU AUG UUC UA	Linear A-RNA	67
5.2[a]	AUG GAU UCU GGU AUG UUC UA	Linear G-RNA	82
V-Shaped Oligonucleotides			
5.3	$A^{2',5'}$(gta tgt)$_{3',5'}$gta tgt	V-DNA	57
5.4	$A^{2',5'}$(GUA UGU)$_{3',5'}$GUA UGU	V-RNA	49
Y-Shaped Oligonucleotides			
5.5	tac ta$A^{2',5'}$(gta tgt)$_{3',5'}$gta tgt	Y-DNA	52
5.6	UAC UAA$^{2',5'}$(GUA UGU)$_{3',5'}$GUA UGU	Y-RNA	6
5.7	UAC UAA$^{2',5'}$(GUA UG*c*)$_{3',5'}$GUA UG*c*	Y-RNA	3
5.8	UAC UAA$^{2',5'}$(GUA UG*ccc*)$_{3',5'}$GUA UG*ccc*	Y-RNA	2.5
5.9	*c*UAC UAA$^{2',5'}$(GUA UG*ccc*)$_{3',5'}$GUA UG*ccc*	Y-RNA	2.5

Notation: small cap letters=deoxynucleotide residues; large cap letters=ribonucleotide residues; *c*=unnatural L-2'-deoxycytidine (L-dC); $^{2',5'}$X=2',5'-phosphodiester linkage; $_{3',5'}$X=3',5'-phosphodiester linkage. IC_{50} values are calculated from the plots of inhibitor concentration versus % RNA splicing (**Figures 5.7-5.8**) and represent the concentration of inhibitor which is capable of hindering the formation of mRNA by 50%. [a]Sequences and results adopted from the thesis of Dr. Ravinderjit Braich (Ref. 212).

Synthetic linear and branched oligonucleotides related to the splicing intermediates derived from *S. cerevisiae* actin pre-mRNA, serve as useful probes to investigate branchpoint recognition in yeast nuclear extract capable of performing pre-mRNA splicing.[208,212] Using our published protocols on bNA synthesis[204,206,233], a variety of potential DNA and RNA substrates with symmetrical 2' and 3' sequences appended to the branchpoint adenosine were synthesized in order to assess the effect of exogenous bNAs on the efficiency of pre-mRNA splicing (**Table 5.1**). The sequences were designed to mirror the absolutely conserved residues present in the BPS (*i.e* UACUAA; A=branchpoint) and the intronic portion of the 5'-splice site (*i.e.* GUAUGU) of a yeast actin pre-mRNA that form the native lariat branched structure (**Figure 5.3**). Since the 2'

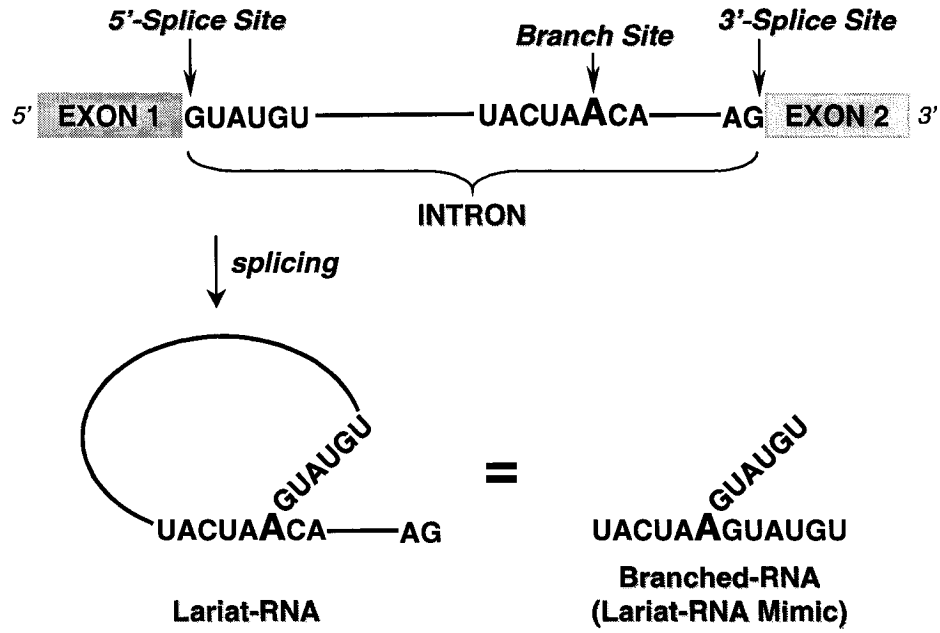


Figure 5.3: Consensus sequences of the yeast actin pre-mRNA transcript used in the *in vitro* splicing assays with *S. Cerevisiae* nuclear extract. Branched RNA is structurally similar to the *in vivo* and *in vitro* synthesized lariat species.

and 3' sequences of the synthetic bNAs are identical, the branch core architecture is slightly different from that of the yeast consensus sequence (*i.e.* 5'..AA$^{2',5'}$(G..)$_{3',5'}$G.. compared to the natural 5'..AA$^{2',5'}$(G..)$_{3',5'}$C..). Nonetheless, mutational analysis has validated that this base alteration may be inconsequential since replacing the conserved 3'-C for a 3'-G in a yeast intron has no discernable effect on splicing.[419] Additionally, single or multiple insertions of unnatural L-dC (see **Table 5.1**) were incorporated at the termini of the bNAs in order to assess their stabilizing effect against the multitude of cellular nucleases[420,421], particularly 3'-exonucleases, present in the yeast extract milieu (compounds **5.7**-**5.9**). Stabilization of the oligonucleotides is an essential criterion since it will increase the amount of bioavailable bNA thereby conceivably giving rise to more pronounced inhibitory profiles. The unnatural terminal residues were introduced during the solid-phase assembly cycle using the commercially available L-2'-deoxycytidine phosphoramidite (ChemGenes Corp.) and required identical activation, coupling and deprotection conditions as the standard DNA phosphoramidite units.[420,421]

The synthetic oligonucleotide substrates were purified by denaturing PAGE or anion-exchange HPLC (see Experimental **Table 7.5**), and extraneous salts removed prior to biological analysis. Sub-optimal gel and HPLC profiles demonstrating the synthesis of bNA **5.6** are shown in **Figure 5.4**. As discussed in Chapter 2, branching of two adjacently tethered oligonucleotide strands with a dilute solution of an adenosine-2',3'-O-bis-phosphoramidite followed by chain extension from the apex of the molecule, results

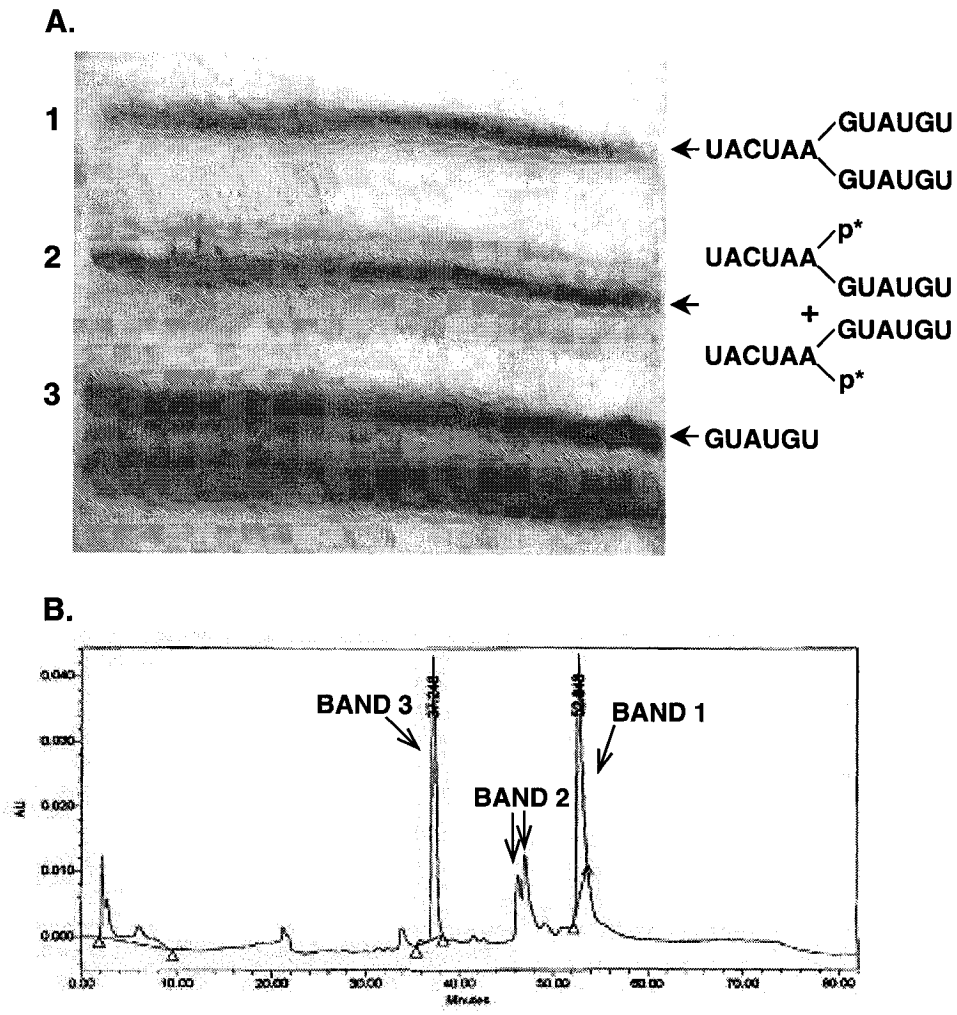


Figure 5.4: PAGE and HPLC analysis of crude **5.6**; UACUAA$^{2',5'}$(GUAUGU)$_{3',5'}$GUAUGU. Panel A: 24% denaturing (7 M urea) preparatory gel of **5.6**. Band 1: full length branched product **5.6** (18-mer); Band 2: unbranched linear isomeric failure sequences; Band 3: truncated hexamer product resulting from the failure of the support bound oligonucleotide to react with the dilute bis-phosphoramidite. Panel B: Anion-exchange HPLC chromatogram of crude **5.6**. Conditions: Buffer A: ddH_2O; Buffer B: 0.2 M $LiClO_4$; Gradient: 0-100% Buffer B over 60 min.; T=50°C.

in the formation of a Y-shaped nucleic acid structure (**Figure 5.4A**; Band 1). Failure of the bis-phosphoramidite to react with one of the two adjacent strands produces a regioisomeric mixture of linear branched failure sequences (**Figure 5.4A**; Band 2). Alternatively, unsuccessful coupling of the bis-phosphoramidite with the support bound precursor strands results in the truncated linear oligonucleotide segments (**Figure 5.4A**; Band 3). Comparatively, the crude mixture was analyzed by anion-exchange HPLC (**Figure 5.4B**) and the identical pattern of product Y-RNA (Band 1), and failure sequences (Bands 2 and 3) were revealed. In fact, resolution of the regioisomeric failure sequences was possible by HPLC, demonstrating an approximately equivalent ratio of the two species. Typically, bNA synthesis was much more efficient than that demonstrated in **Figure 5.4** and this example was simply utilized to demonstrate the main abbreviated products which can result from ineffective coupling of the bis-phosphoramidite with the support-bound oligonucleotide chains. Prior to assessing the inhibitory potentials of the individual bNAs in a yeast splicing system, the stabilities of the compounds against cellular nucleases had to be gauged in the presence of the nuclear extract under typical splicing conditions.

5.2.2.1. Nuclease Sensitivity of Branched Oligonucleotides Containing Unnatural L-dC Termini in Yeast (*S. Cerevisiae*) Nuclear Extract

Radioisotopic (*i.e.* ^{32}P) 5'-end labeling studies on synthetic linear and branched oligonucleotides previously studied for their splicing inhibitory potential by Dr. Ravi Braich of our research group have demonstrated substantial degradation of the substrates, seemingly by exonucleases present in the yeast extract milieu (**Table 5.1**; compounds **5.1**-**5.2** and **5.3**-**5.6**).[212] The incorporation of unnatural L-2'-deoxynucleotide residues at the ends of an otherwise unmodified D-oligonucleotide have been shown to provide adequate stability towards exonuclease hydrolysis.[420,421] As such, the nuclease sensitivity of Y-RNA **5.6** containing natural terminal ribonucleotide residues was comparatively evaluated alongside similar Y-RNAs containing suitably stabilizing L-dC terminal units (**5.7**-**5.9**). The remaining portions of the molecules enclosed the identical BPS and 5'-intronic splice

site segment as the native actin lariat intron (**Table 5.1 and Figure 5.3**). The 5'-termini of the Y-RNAs (**5.6**-**5.9**) were radiolabeled using [γ-^{32}P]-ATP and T4 polynucleotide kinase (T4 PNK) as described in section 7.7.1. Introduction of the label in compound **5.9** necessitated a slightly longer reaction time (*ca.* 2 h) owing to the presence of an unnatural L-dC at the 5'-terminus. It has been reported that T4 PNK tolerates even non-nucleotidic modifications at the 5'-terminus of oligonucleotides[422], and demonstrates a lack of stereodifferentiation between D- and L-nucleotides and thus a limited specificity for only natural enantiomeric forms of nucleotide residues.[423] Nonetheless, an extended reaction time appeared to improve the amount of radiolabeled oligonucleotide obtained (data not shown). Substrate Y-RNAs **5.6**-**5.9** were incubated with diluted yeast nuclear extract (20× dilution) for variable lengths of time (15-60 min). Potassium phosphate (5 mM final concentration of KH_2PO4) was added as a general phosphatase inhibitor to avoid potential cleavage of the 5'-monophosphate esters by nonspecific enzymatic dephosphorylating activities. Degradation profiles were monitored by partitioning the oligonucleotide mixture on a denaturing gel (16%, 7M urea), and the cleavage products visualized by autoradiography (**Figure 5.5**). A qualitative assessment of the autoradiogram clearly established that the bNAs containing modified L-dC termini (**5.7**-**5.9**; lanes 6-17) were more resistant to enzymatic digestion than the bNA containing the natural ribose terminus (**5.6**; lanes 2-5), as judged by the disappearance of initial radiolabeled RNA and the appearance of shorter cleavage products. The extent of nuclease hydrolysis was further confirmed by quantitating the amount of residual bRNA and dividing by the total radioactivity in each lane. The obtained values were then normalized with respect to the bNA lacking any extract (t=0). Under the conditions employed, bNA **5.6** was degraded by 50% after only 15 minutes incubation (**Figure 5.5**). Alternatively, bNAs **5.7**-**5.9** were almost completely resistant to nuclease digestion even after exposure to the extract for 1 hour (**Figure 5.5**). Additionally, it did not appear that incorporating more than one L-dC at the 3'-terminus (compounds **5.8** and **5.9**) increased the robustness of the molecules, since one inclusion demonstrated adequate stability (**Figure 5.5**; compare **5.7**-**5.9**). It also seemed that 5'-exonuclease activity was not a serious issue since those substrates that had natural ribose 5'-ends (**5.7**-**5.8**) were just as stable as the 5'-L-dC containing bNA (**5.9**).

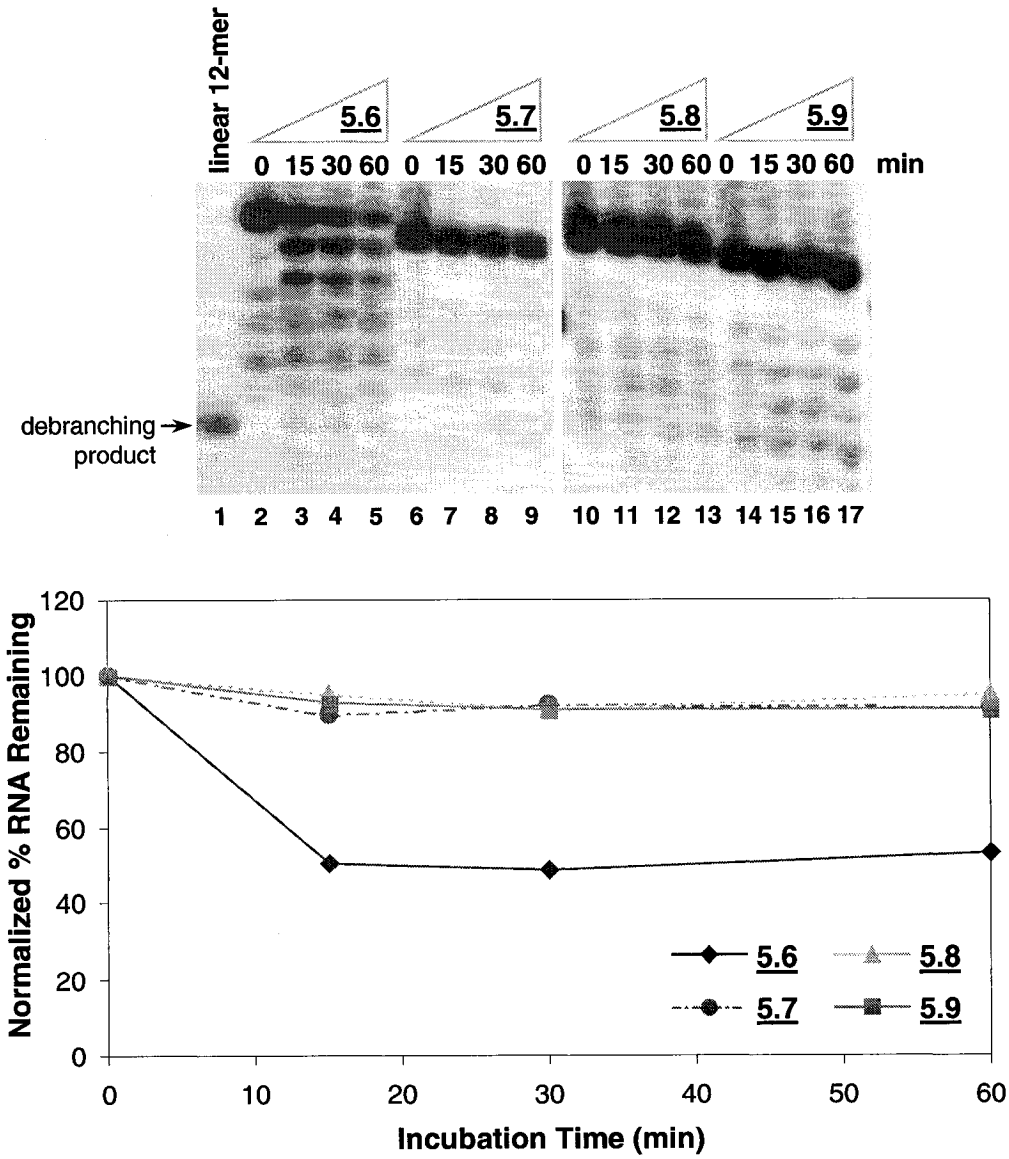


Figure 5.5: Effect of unnatural L-dC termini on the nuclease stability of Y-RNAs in yeast nuclear extract. Y-RNAs were incubated with a 20-fold dilution of yeast extract in 100 mM Tris-HCl, pH 7.8. 50 mM KH_2PO_4 was added as a general phosphatase inhibitor. The % RNA remaining was determined by quantitation of the precursor Y-RNA bands and the total radioactivity in each lane in the autoradiogram (above) using the UN-SCAN-IT program (Silk Scientific) and normalizing with respect to the t=0 time point (100 % RNA). Linear RNA 12-mer (debranching product): 5'-UAC UAA GUA UGU-3'.

Since yeast nuclear extract is a typical source of RNA debranching activity[116], it was also expected that one of the main degradation products for each of the substrate bNAs would be the corresponding linear RNA lacking the 2'-branch extension.[115]

Accordingly, a 12-nucleotide RNA sequence (*i.e.* 5'-UACUAAGUAUGU-3') corresponding to the debranching product of **5.6** was run alongside the bNAs in the enzymatic digestion assay. Surprisingly, the nuclear extract appeared to be devoid of debranching activity at the concentration ultilized. Whether this result is due to the concentration of extract used or the method employed for preparing the extract itself is uncertain at this time.

5.2.2.2. Inhibition of Yeast Actin pre-mRNA Splicing with Branched RNAs

A viable yeast actin pre-mRNA transcript was prepared by run-off T7 RNA polymerase transcription of an EcoR1 linearized pGEM®9Zf(-) DNA vector containing the actin gene as template (see section 7.12.1.3).[415] Internal radiolabels (^{32}P) were introduced throughout the transcript by way of [α-^{32}P]-UTP. *In vitro* splicing reactions contained 30% v/v yeast nuclear extract and were performed according to the splicing method developed by Abelson.[424] Assays were performed at 37°C and the actin-containing pre-mRNA splicing reactions were supplemented with 1-100 μM of cold V-shaped or Y-shaped oligonucleotide inhibitor. The protein content of the splicing mixtures was removed by digestion with Proteinase K followed by phenol/chloroform extraction. By using a short reaction time (*ca.* 20 min), the actin pre-mRNA, lariat-containing intermediates, and mature mRNA were readily separated and identified by denaturing PAGE (Schematic representation **Figure 5.2** and **Figure 5.6**). Additionally, the effects of the distinctive bNA inhibitors on the overall splicing efficiency were evaluated by quantitating the amount of mature actin RNA (mRNA) produced as a percentage of the total radioactivity in each lane, and normalized with respect to the reaction lacking any bNA (0 μM). This method of quantitation, however, does not yield information about which chemical step in the splicing reaction (Step I or II) is being affected by the bNA inhibitor. Alternatively, obstruction of step one or step two of the splicing reaction could also be monitored by measuring either; (1) a decrease in the amount of lariat-3'-exon intermediate (Step I inhibition), or (2) an accumulation in the amount of lariat 3'-exon intermediate (Step II inhibition) as shown in **Figure 5.1**. However, since a clearly

distinct band for the lariat-3'-exon intermediate was not always apparent, mRNA quantitation was exploited instead.

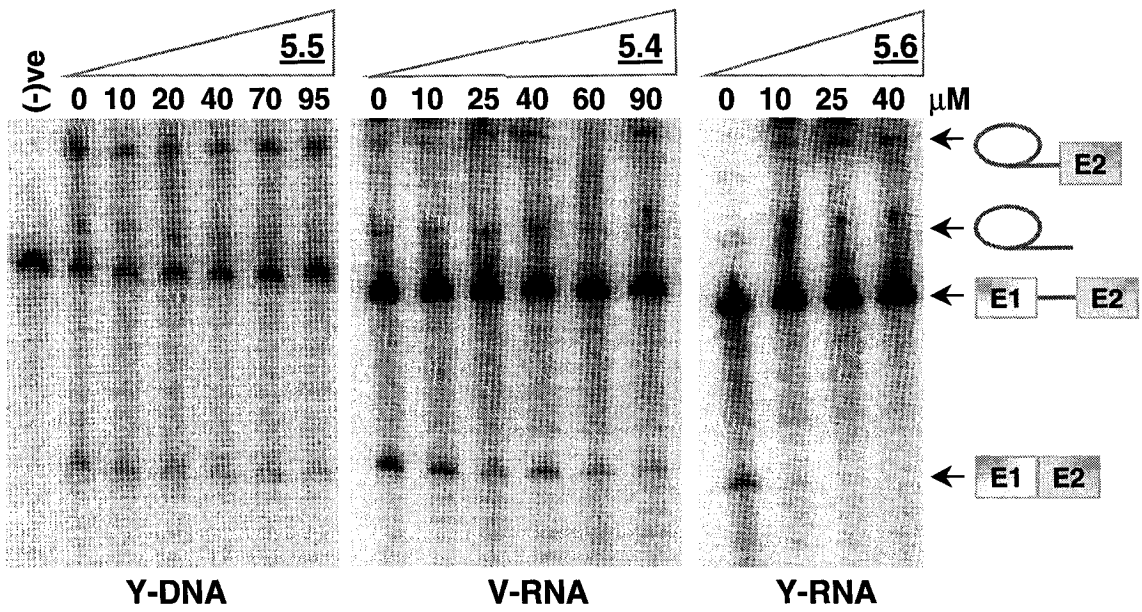


Figure 5.6: *In vitro* splicing inhibition profile of an internally radiolabeled yeast actin pre-mRNA substrate in the presence of increasing concentrations of cold V-RNA (**5.4**), Y-DNA (**5.5**) and Y-RNA (**5.6**). Splicing products and intermediates were partitioned on a 5% denaturing (7 M urea) PAGE. The gel products were visualized using a Molecular Dynamics Phosphorimager.

Previously[208,212], it has been suggested that the most potent inhibitory profiles were observed with bNAs containing a well defined branchpoint structure consisting of 5'-, 3'- and 2'-appendages off the branchpoint adenosine. Although moderate suppression was also observed with Y-shaped oligothymidylates[208], naturally, the most potent splicing inhibition was observed with a Y-shaped RNA molecule of identical sequence composition to **5.6**.[212] Such structures most closely mimicked the naturally occurring lariat structure. Additionally, high concentrations (90-100 μM) of linear RNA oligonucleotides were also capable of inhibiting the formation of the spliced mRNA product.[212] A similar trend was also observed with linear oligothymidylates and appeared to represent non-specific effects, however more studies were needed to confirm this. Since many of the stated results were only obtained with one trial, it seemed

imperative to repeat some of the inhibition experiments on substrates previously reported in order to determine the importance of the branched topology and sugar (*i.e.* deoxyribose or ribose) on specific nature of recognition of the exogenous substrates (compounds **5.1**-**5.6**). Additionally the inhibitory profiles of those bNAs containing stabilizing L-dC termini were also evaluated (compounds **5.7**-**5.9**). Such studies were conducted in collaboration with Dorian Anglin and Dr. James L. Friesen of the University of Toronto (Banting and Best Department of Medical Research).

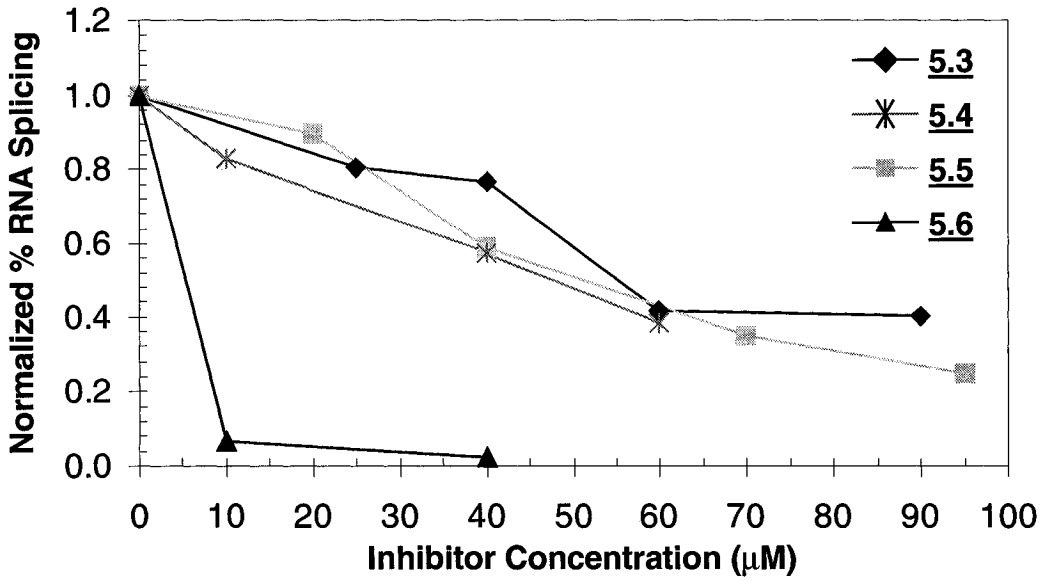


Figure 5.7: Comparison of the yeast splicing inhibition profiles of branched DNA and RNA molecules (V-DNA & V-RNA; Y-DNA and Y-RNA). The normalized % RNA splicing of the actin transcript was calculated as described previously.

As demonstrated in **Figures 5.6-5.8**, the exogenous DNA and RNA molecules each affect the efficiency of splicing, yet to different degrees as demonstrated by the decrease in the amount of mRNA formed in the individual assays. Specifically, bRNA (**5.4** & **5.6**) molecules were dominant inhibitors of the splicing reaction compared to their homologous bDNA (**5.3** & **5.5**) counterparts (**Figure 5.7**). Most notably however, was the potent inhibitory profile of the Y-RNA **5.6** which displayed a markedly enhanced IC_{50} of 6 μM at the concentration range studied (10-40 μM). The synthetic Y-shaped RNA **5.6** inhibited the formation of the mRNA product to a much greater extent than the Y-DNA **5.5** with a near 97% reduction in splicing efficiency at 10 μM compared to 75% at

95 μM (**Table 5.1**). Similarly, the V-RNA **5.4** (IC_{50}=49 μM) displayed a more effective inhibitory activity than its V-DNA **5.3** correlative (IC_{50}=57 μM). This significant observation suggests that branch recognition is highly dependent on either the 2'-hydroxyl, which is present in the ribose sugar, or on the adopted sugar conformation of the bNA which is expected to be C2'-*endo* for the bDNA **5.5** and C3'-*endo* for the bRNA **5.6**. It is well established that within the spliceosome there exist a number of RNA-binding proteins bearing one or more RNA-recognition motifs.[88] Conceivably, one or more of these factors may display differential affinities for RNA over DNA oligonucleotides. This is not surprising since previous work has shown that incorporation of modified nucleotides, particularly deoxynucleotides at either the 3'-splice site[425] or branch point[69] results in deleterious effects on the overall splicing reaction. A 3'-splice site mutation dramatically altered the rate of the second step of splicing, whereas a deoxyadenosine residue in the place of the conserved riboadenosine at the branchpoint resulted in a substantial decrease to step two, as well as cryptic branch site formation at the G-nucleotide directly adjacent to it. When the adjoining G-nucleotide was also replaced with a 2'-deoxynucleotide substitution, a significant block to the first step was also apparent without utilization of any other immediately neighboring nucleotides.[69] Additionally, antisense oligonucleotides targeted towards the branchsite recognition sequences of the U2-snRNA (positions 29-43) have demonstrated a weaker binding affinity for oligodeoxyribonucleotides (DNA) compared with oligoribonucleotides of identical sequence composition.[426] These results support the notion that the 2'-hydroxyl units and/or the backbone sugar conformation enclosed within RNA oligonucleotides are requisite signals for proper spliceosomal recognition by various factors and effective splicing chemistry.

Directly comparing the potencies of Y-RNA **5.6** versus the V-shaped oligonucleotide **5.4**, another distinct observation becomes obvious. The fully formed Y-RNA exhibits a noticeable 8-fold enhancement in inhibitory activity compared to the V-RNA that lacks the 5'-extension (**Figure 5.8** and **Table 5.1**). Clearly, this suggests that recognition of the exogenous oligonucleotide structures not only involves the branch core structure itself,

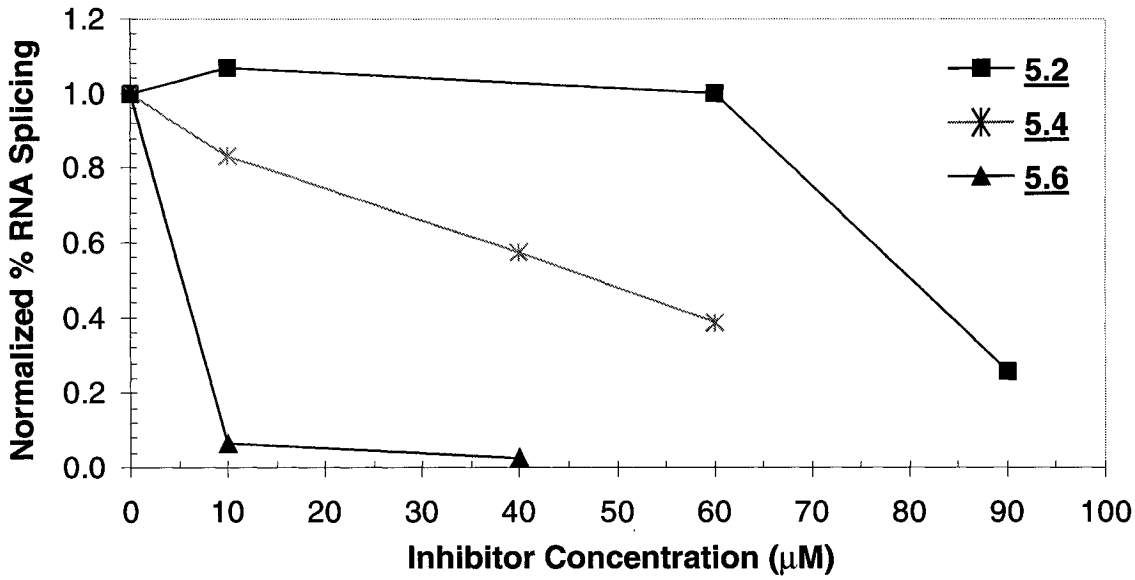


Figure 5.8: Comparison of the yeast splicing inhibition profiles of linear, V- and Y-RNA. The % RNA splicing of the actin transcript was calculated by densitometric analysis of mRNA product band divided by the total radioactivity in the respective lane (Molecular Dynamics Image-Quant® software). The individual values were normalized with respect to the lane lacking any inhibitor (0 μM).

but the sequences upstream of the branchsite as well. A genetic screen applied to mutants in the invariably conserved UACUAAC branch region in yeast has demonstrated that all the nucleotide positions preceding the branchpoint adenosine are important for intron recognition.[427] Additionally, a particularly important protein, Mud2p, crosslinks to the nucleotides within this domain, and the results indicate that this protein factor is required for efficient intron recognition by the splicing machinery. Mud2p is the yeast ortholog of the mammalian splicing factor U2AF^{65}. Alternatively, the disparity in activity demonstrated by the V- and Y-shaped molecules might also suggest that the three dimensional structure of the inhibitor is a valuable recognition motif. Specifically, conformational and NMR studies on branched tri- and tetranucleotides have revealed that the branchpoint adenine base is heavily involved in an intramolecular base-stacking interaction between the G-nucleotide at the 2'-position and the purine moiety at the 5'-position off the branch.[314,315] Furthermore, the ribose sugar of the branchpoint adenosine appears to adopt a preferred C2'-*endo* sugar pucker in such structures. Such a

localized effect is speculated to serve as a point of distortion in the normal A-type RNA helix[315], and as such, may act as a signal for recognition by a putative branch recognition factor. Overall, the repeated findings corroborate the fact that branched oligonucleotides, especially Y-RNAs of defined base composition are potent inhibitors of the pre-mRNA splicing pathway. Nonetheless, the most active bNA, **5.6** consisted of highly vulnerable RNA constituents at the 3'-terminus, and likely limit the amount of biologically accessible material. As such, the study merited the evaluation of the more robust L-2'-deoxycytidine containing structures.

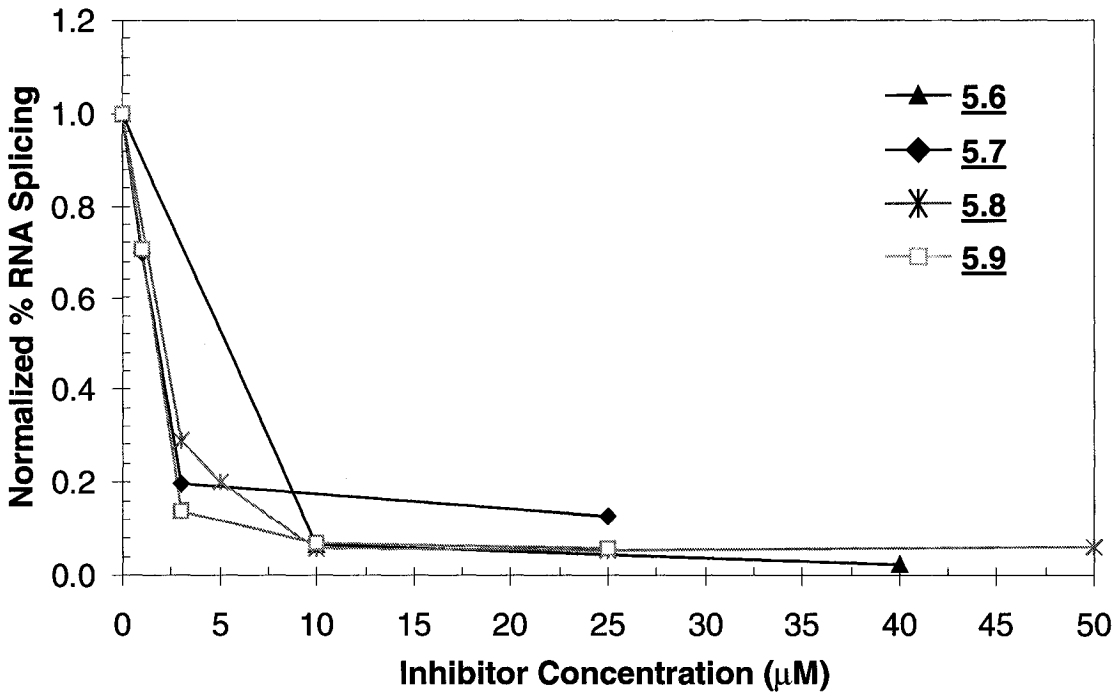


Figure 5.9: Comparison of the yeast splicing inhibition profiles of Y-RNA incorporating different numbers of terminal unnatural L-dC residues. The % RNA splicing of the actin transcript was calculated as described previously.

When cold Y-RNA molecules **5.7**-**5.9** were supplemented in the pre-mRNA splicing reaction, a similar inhibitory profile was observed to that seen with bNA **5.6**, which lacked the 3'- and/or 5'-L-dC stabilizing residues (**Figure 5.9**). Lower concentrations of inhibitor (1, 3 and 5 μM) were utilized for these substrates in order to determine more accurate IC_{50} values. In all three cases, the IC_{50}'s were approximately in the 2-3 μM range. In fact, at a concentration of 10 μM of exogenous bNA, splicing (*i.e.* mRNA

formation) was almost completely abolished. The inhibitory reproducibility among the bNAs indicates that the unnatural L-dC termini are well tolerated, and are likely too far away from the branch recognition region of the molecule to hamper sequestration of the putative branch recognition factor. The global results indicate that the most vital recognition domains in the bNA are the branch core structure itself and the 5'-extension off the branch (*i.e.* sequences preceding the branchpoint adenosine).

Whether splicing inhibition is the result of branch recognition prior to step one or step two remains undetermined at this time given that a the behavior of the lariat-3'-exon intermediate with increasing concentration of inhibitor could not be thoroughly assessed. It seems reasonable to speculate that branch recognition would only occur after the first chemical step of the splicing reaction, since this would result in the formation of a branched lariat-RNA intermediate species, which shares distinct similarities with the most potent inhibitor, the Y-RNA. Alternatively, should inhibition occur early in the splicing process (*i.e.* prior to step one), this would indicate that branch recognition is also an early event effected by a factor which is present in the initial spliceosomal complex and remains bound to the branchsite region after step one. Remarkably, the U2-snRNA is one of those non-protein spliceosomal factors which has been shown to remain tethered to the intronic branchsite after lariat-3'-exon formation[111], and its key role in early branchsite recognition is already well established.

5.2.3. Inhibition of *In Vitro* pre-mRNA Splicing in Mammalian Nuclear Extract (HeLa cells) by Branched RNAs

As mentioned previously, the ability to study pre-mRNA splicing reactions *in vitro* has been realized with both yeast and human (HeLa) extracts. Given that the entire sequence of the *S. cerevisiae* yeast genome has been known since 1996[428], the benefit of using yeast for the investigation of RNA splicing has been predominantly the ability to perform genetic manipulations of the proteins and snRNAs to gain insights into the interactions of the spliceosome. On the other hand, consistency in terms of batch-to-batch activity of the

extracts for *in vitro* assays, are very difficult to reproduce and typically require much optimization.[429] This irreproducibility was also noted in our own extract preparations. Alternatively, human extracts have allowed for much dissection of purified splicing complexes and components, and published protocols for their preparation typically yield reproducibly active extracts.[92,416,417] It has been shown, however, that mammalian *in vitro* splicing reactions are heavily dependent on the concentration of potassium chloride (KCl) in the splicing milieu, with an optimum concentration in the 40-60 mM range. As such, HeLa nuclear extract for *in vitro* splicing inhibition assays by exogenous bNAs were prepared according the well-cited method of Dignam[430], and were a generous donation from Dr. Andrew MacMillan (University of Alberta). Nuclear extracts were used as received, stored at -80°C and thawed on ice prior to use.

A variety of V-shaped and Y-shaped DNA and RNA synthetic inhibitors were synthesized according to the convergent bNA methodology described previously (**Table 5.2**).[204,206,233] In the case of compounds **5.19** and **5.20**, the arabino-adenosine bis-phosphoramidite was prepared by Dr. Katya Viazovkina of our research group in an identical fashion to the ribo-adenosine bis-phosphoramidite (section 7.8.5), however, the starting precursor was the arabino-adenosine nucleoside.[187] The coupling conditions and times during solid-phase synthesis were identical to those employed for bRNA synthesis. Some of the sequences had been previously utilized to study branchpoint recognition during *in vitro* splicing in yeast (compounds **5.6**-**5.9**). Additional bDNA and bRNA sequences composed of terminal L-dC residues, 2'-O-methyl pyrimidine substitutions and a non-native arabino-adenosine branchpoint moiety were also devised to assess their nuclease sensitivity during *in vitro* splicing as well as to determine their inhibitory potential (compounds **5.14**-**5.20**). Furthermore, linear DNA and RNA molecules (compounds **5.10**-**5.13**) were also synthesized to help elucidate the affinity of factor-binding to such sequences compared to bNAs of similar nucleotide composition. The conserved yeast lariat intronic sequence was maintained for the studies in mammalian extract since the UACUAAC branch region has been shown to be the optimal branch site recognition motif in both systems.[57] This branch-site sequence is known to form base

Table 5.2: List of linear, V- and Y-shaped oligonucleotides utilized as inhibitors of *in vitro* pre-mRNA splicing in mammalian nuclear extract (HeLa)

Code	Sequence (5'→3')	Topology
Linear Oligonucleotides		
5.10	tac taa gta tg*c*	Linear DNA
5.11	UAC UAA GUA UGU	Linear RNA
5.12	UAC UAA GUA UC*c*	Linear RNA
5.13	**UAC UAA GUA UG***c*	Linear RNA
V-Shaped Oligonucleotides		
5.14	$A^{2',5'}$(gta tg*c*)$_{3',5'}$gta tg*c*	V-DNA
5.15	$A^{2',5'}$(GUA UG*c*)$_{3',5'}$GUA UG*c*	V-RNA
Y-Shaped Oligonucleotides		
5.16	tac ta$A^{2',5'}$(gta tg*c*)$_{3',5'}$gta tg*c*	Y-DNA
5.6	UAC UA$A^{2',5'}$(GUA UGU)$_{3',5'}$GUA UGU	Y-RNA
5.7	UAC UA$A^{2',5'}$(GUA UG*c*)$_{3',5'}$GUA UG*c*	Y-RNA
5.8	UAC UA$A^{2',5'}$(GUA UG*ccc*)$_{3',5'}$GUA UG*ccc*	Y-RNA
5.9	*c*UAC UA$A^{2',5'}$(GUA UG*ccc*)$_{3',5'}$GUA UG*ccc*	Y-RNA
5.17	*ccc*UAC UA$A^{2',5'}$(GUA UG*ccc*)$_{3',5'}$GUA UG*ccc*	Y-RNA
5.18	***ccc*UAC UA**$A^{2',5'}$**(GUA UG*ccc*)**$_{3',5'}$**GUA UG*ccc***	Y-RNA
5.19	UAC UA(aA)$^{2',5'}$(GUA UGU)$_{3',5'}$GUA UGU	(ara A)-Y-RNA
5.20	UAC UA(aA)$^{2',5'}$(GUA UG*c*)$_{3',5'}$GUA UG*c*	(ara A)-Y-RNA

Notation: small cap letters=deoxynucleotide residues; large cap letters=ribonucleotide residues; large cap bold letters=2'-OMe-ribonucleotide residues *c*=unnatural L-2'-deoxycytidine (L-dC); aA=arabino-adenosine branchpoint; $^{2',5'}$X=2',5'-phosphodiester linkage; $_{3',5'}$X=3',5'-phosphodiester linkage.

pairs with the complementary sequence GUAGUA in U2 small nuclear RNA (snRNA) during mRNA splicing in the yeast (*S. cerevisiae*), thereby forcing the branchpoint adenosine nucleotide to bulge out of the helix.[66] Even though mammalian branch sites conform only weakly to the YURAC consensus (Y=pyrimidines; R=purine), the GUAGUA element is preserved in mammalian U2-snRNA. The linear, V-shaped and Y-shaped constructs were purified by either denaturing PAGE or anion-exchange HPLC and the extraneous salts removed prior to biological testing (see **Table 7.5** of Experimental).

The nucleotide composition of the individual samples was confirmed by negative-mode MALDI-TOF-MS and/or comparison to oligonucleotides standards of similar sequence constitution and length.

5.2.3.1. Nuclease Sensitivity of Linear and Branched Oligonucleotides in Mammalian (HeLa) Nuclear Extract

The nuclease resistance profiles of select linear (**5.11**-**5.13**) and Y-shaped (**5.6**-**5.9** and **5.17**-**5.20**) RNAs were conducted under splicing conditions (*i.e.* 25% extract, T=30°C) for variable lengths of time (t=0-90 min). Substrate oligonucleotides were terminally labeled at their 5'-positions with γ-[^{32}P]-ATP by way of bacteriophage T4 PNK (Experimental section 7.7.1) prior to the assay. Potassium phosphate (KH_2PO_4) was once again employed as a general phosphatase inhibitor to avert significant loss of the radioisotopic signal by ubiquitous phosphatases in the extract during incubation. After incubation, the protein content of the mixture was removed by standard phenol/chloroform extraction. Since the oligonucleotide products were resolved on a denaturing gel (16%, 7M ureas), this cleanup prevented the oligonucleotides from becoming stuck in the wells. The amount of RNA degradation was ascertained by quantifying the amount of residual RNA at each time point, dividing by the total radioactivity in the lane and normalizing with respect to the RNA lacking any extract (t=0 min).

The degradation profiles of the various substrate RNAs revealed some interesting trends. As described for the yeast system, one insertion of an unnatural L-dC residue at the 3'-terminus of a bRNA was sufficient to provide adequate stability against the cellular exonucleases present in the extract (section 5.2.1.1). When linear RNA **5.11** (no 3'-L-dC residue) was incubated with the extract, a significant amount of degradation was obvious, even after 15 min (**Figure 5.10A** and **5.11**; lanes 1-5) resulting in almost complete loss of the signal after the full 90 min. Alternatively, **5.12**, a linear RNA bearing almost identical sequence composition, however incorporating one L-dC unit at the 3'-end, was highly resistant to the nuclease activity present in the extract (**Figure 5.10A** and **5.11**; lanes 11-

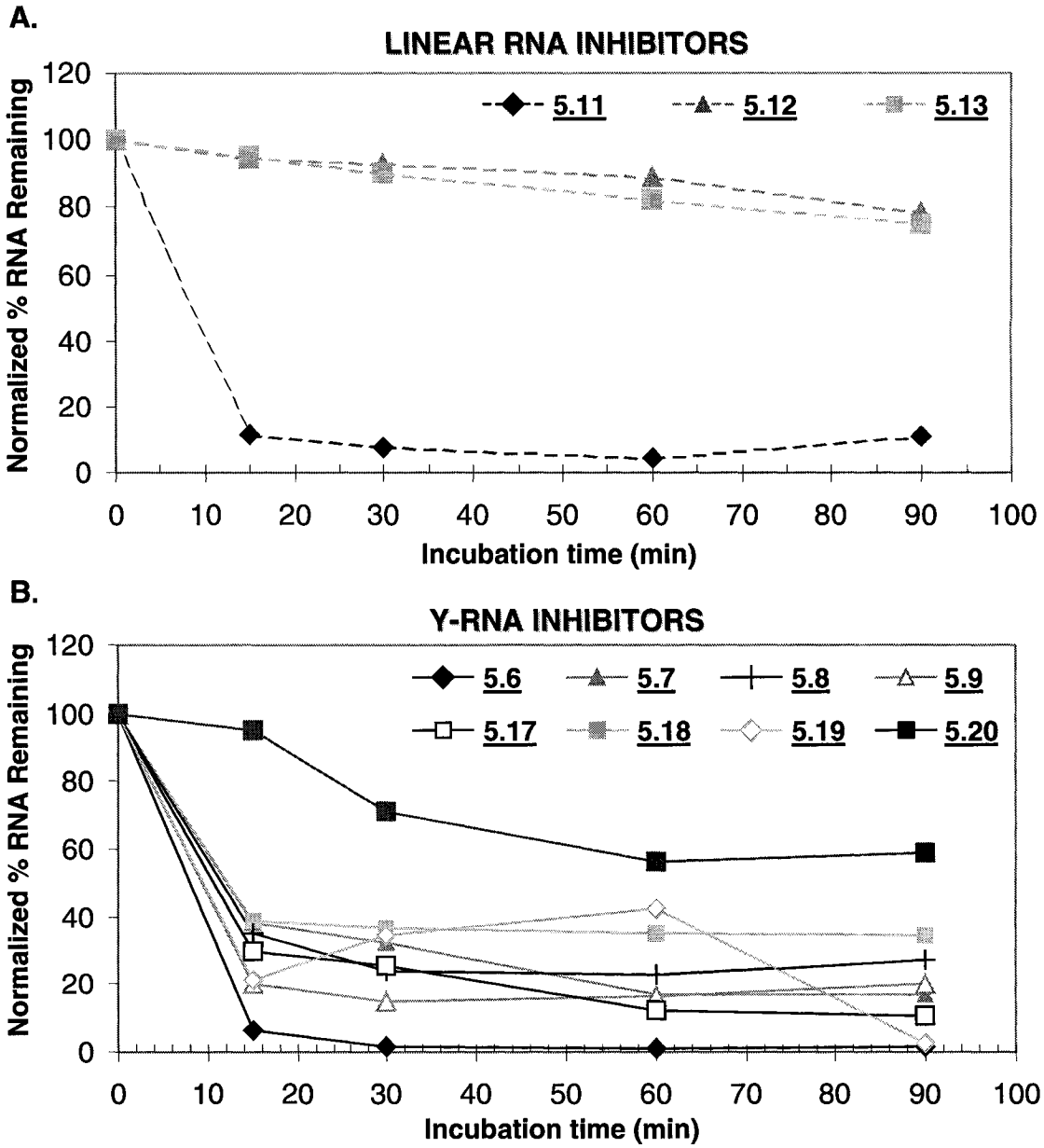


Figure 5.10: Charts demonstrating the nuclease stability of various linear RNAs (**5.11**-**5.13**) and Y-RNAs (**5.6**-**5.9** and **5.17**-**5.20**) under mammalian *in vitro* splicing conditions. Panel A: nuclease sensistivity of linear RNAs **5.11**-**5.13** in the presence of HeLa extract. Panel B: nuclease sensitivity of Y-RNAs **5.6**-**5.9** and **5.17**-**5.20** in the presence of HeLa extract. The % RNA remaining was determined by quantitation of the precursor Y-RNA bands divided by the total radioactivity in each lane in the respective autoradiograms using the UN-SCAN-IT program (Silk Scientific) and normalizing with respect to the t=0 time point (100 % RNA).

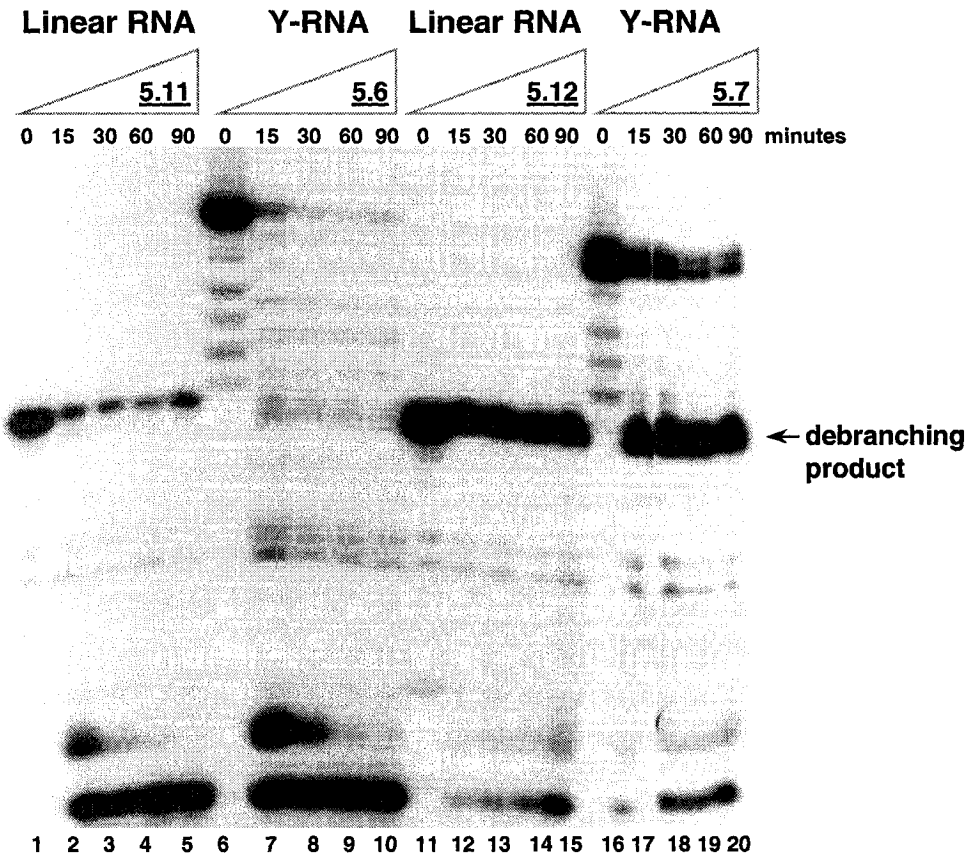


Figure 5.11: Effect of unnatural L-dC termini on the nuclease stability of linear and Y-RNAs in HeLa nuclear extract. Y-RNAs were incubated with a 25% HeLa extract under splicing conditions. 50 mM KH_2PO_4 was added as a general phosphatase inhibitor (5 mM final concentration). Samples were resolved on a 16% denaturing gel (7M urea) and the bands visualized by autoradiography.

15). Similarly, the bNA **5.6** was completely degraded after 90 min (**Figure 5.10B** and **5.11**; lanes 6-10), whereas the 3'-stabilized compound 5.7 demonstrated considerable resistance (**Figure 5.10B** and **5.11**; lanes 16-20). Nonetheless, the amount of residual RNA in the more nuclease resistant **5.7** was still decreasing with time, indicating that a competing hydrolysis reaction was likely the culprit (**Figure 5.11**; lanes 16-20). As stated formerly, yeast and HeLa extract preparations typically contain the 2'-phosphodiesterase RNA debranching activity.[109] In fact, this specific 2'-branch cleaving activity found in HeLa extracts has been suitably exploited in the characterization of the DNA and RNA lariat structures previously synthesized in sections 2.4.3 and 3.3.5. As such, 2'-debranching of the bRNA substrate **5.7** would result in the formation of a linear

RNA oligonucleotide sequence identical to **5.12** (*i.e.* Y-RNA lacking the 2'-extension). Indeed, the accumulation of one main hydrolysis product was clearly evident, and its migratory rate was consistent to that of the linear debranching product **5.12** (**Figure 5.11**; compare lanes 11-15 to 16-20). In fact, the nuclease resistance of this product dodecanucleotide is also undoubtedly seen, suggesting that although the 3'-termini of the bNA **5.7** are extremely resistant to cleavage, the disposition of the native branchpoint is a recognition element for hydrolysis by the HeLa debranching enzyme. Quantitation of the remaining RNA indicates that approximately 60% of the Y-RNA **5.7** is debranched after only 15 min (**Figure 5.10B**). This however, was not the case for bNA **5.6**. Seeing as debranching of **5.6** produces the linear construct **5.11**, degradation of this debranched product by exonucleases is still conducive given that it lacks any stabilizing 3'-termini. This is established by the absence of an accumulated debranched product band in the degradation profile of **5.6** (**Figure 5.11**; lanes 6-10). Similarly, other 5'- and/or 3'-L-dC stabilized bRNAs (**5.8**-**5.9** and **5.17**) demonstrated substantial resistance (compared to **5.6**) to the exonuclease activity present in the nuclear surroundings (**Figure 5.11**), however, debranching of the 2'-extension and accumulation of the nuclease resistant linear products were also noteworthy (gel autoradiogram not shown). The similar degradation profiles exuded by the terminally modified bNAs (**5.7**-**5.9** and **5.17**) indicates that multiple L-dC inclusions at the 3'-terminus or at the 5'-terminus are unnecessary, establishing that one insert of L-dC at the 3'-end is enough to impart adequate nuclease resistance. Once again, the 5'-terminal modifications (**5.9** and **5.17**) did not appear to induce greater stability of the structures, suggesting that a negligible amount of 5'-exonucleases were present in the extract mixture. Additionally, some loss of RNA signal in all the substrates may also be due to endonuclease activity as well as hydrolysis of the radioactive 5'-phosphate monoester probe by dephosphorylases in the extract.

A second nuclease resistant modification, 2'-O-methyl-ribonucleoside (*i.e.* 2'-OMe), was also investigated in both linear and Y-shaped RNA constructs (compounds **5.13** and **5.18**). Specifically, the ribopyrimidines (C and U) found in sequences **5.12** and **5.17** were replaced by 2'-OMe ribopyrimidine units. Our rationale for this pyrimidine-nucleotide modification stems from previous research conducted on the synthesis and stability

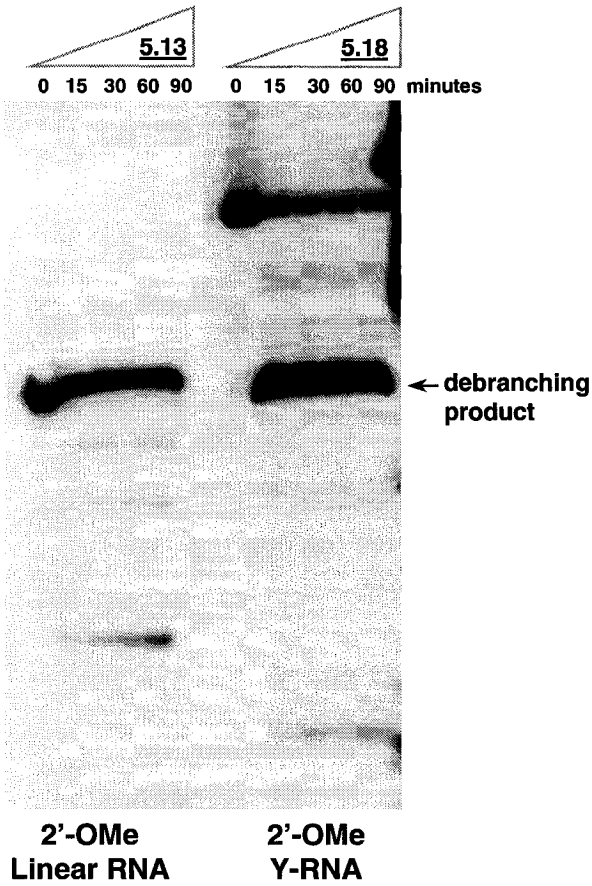


Figure 5.12: Effect of incorporating 2'-O-methyl pyrimidine units and unnatural L-dC termini on the nuclease stability of Y-RNAs in HeLa nuclear extract. Y-RNAs were incubated with a 25% (v/v) HeLa extract under splicing conditions. 50 mM KH_2PO_4 was added as a general phosphatase inhibitor. Samples were resolved on a 16% denaturing gel (7M urea) and the bands visualized by autoradiography.

profiles of synthetic ribozymes, in which alteration of the natural RNA pyrimidines to 2'-OMe substituents greatly enhanced the robustness of the molecules in biological media.[431] Additionally, such modifications were expected to have the same predominant sugar conformation as the native RNA constituents, namely C3'-*endo*.[432,433] The degradation profile of the linear RNA oligonucleotide containing 2'-OMe-pyrimidine ribonucleotide inserts (**5.13**) indicated that the compound was indeed nuclease resistant, however to the same degree as **5.12** which merely contained one 3'-L-dC unit (**Figures 5.10A** and **5.12**). Alternatively, the 2'-OMe modified bNA **5.18** displayed a comparable nuclease stability to the 3'-L-dC containing bNAs (**5.7**, **5.8** and **5.14**) at shorter incubation

times (15 and 30 min), however at longer reaction times, the amount of residual RNA tapered off rather than degrading further (**Figure 5.10B** and **5.12**). This may suggest a stabilizing role for these constituents against endonucleases in the cellular extract. Analogous to the L-dC modified bNAs, debranching of the 2'-OMe modified bNA **5.18** was also clearly manifest as established by the accumulation of one main product band with a migratory rate similar to that of the linear 2'-OMe construct **5.13**. Although the degradation profiles of these 2'-modified substrates were nearly identical to the 3'- and 5'-L-dC containing ribonucleotides, their potential to inhibit pre-mRNA splicing in HeLa extract was still warranted, given that such inclusions boast the same sugar conformation as the best inhibitors in the yeast system, the Y-RNAs.

Although ample protection is imparted by including a minimum of one terminal 3'-L-dC constituent into an otherwise unmodified bRNA structure, the fact remained that the 2'-extension of the bNAs were highly vulnerable to the debranching activity present in the HeLa extract milieu. Incorporation of a branchpoint modification that resists the 2'-scissile activity of the debranching enzyme would therefore increase the effective concentration of branch inhibitor in the pre-mRNA splicing assays. Recent work emerging from our research group has demonstrated that branched RNAs of identical sequence constitution as **5.6**, however, incorporating the 2'- and 3'-epimers of riboadenosine at the branchpoint (*i.e.* arabinoadenosine (ara-A); **5.19** and xyloadenosine), are not substrates for the yeast RNA lariat debranching enzyme (yDBR).[213] Our preliminary results indicate that only ribose as a branchpoint directs optimal debranching activity, suggesting that the configuration of vicinal phosphate groups and/or sugar pucker of the branchpoint to be a crucial element for optimal enzyme recognition and hydrolytic efficacy. As such, compounds **5.19** and **5.20** containing an ara-A branchpoint, were synthesized to examine their susceptibility to either or both exonuclease hydrolysis and the HeLa RNA debranching activity (**Table 5.2**). Upon incubation with HeLa extract, bNA **5.19**, which lacks an L-dC modified 3'-terminus, is rapidly degraded (**Figure 5.10B** and **Figure 5.13**; lanes 3-7). Conversely, the L-dC-containing ara-A Y-RNA (**5.20**) is degraded much more slowly (**Figure 5.10B** and **Figure 5.13**; lanes 8-12). Most notably however, is the absence of a 12-nucleotide long 3'-L-dC stabilized

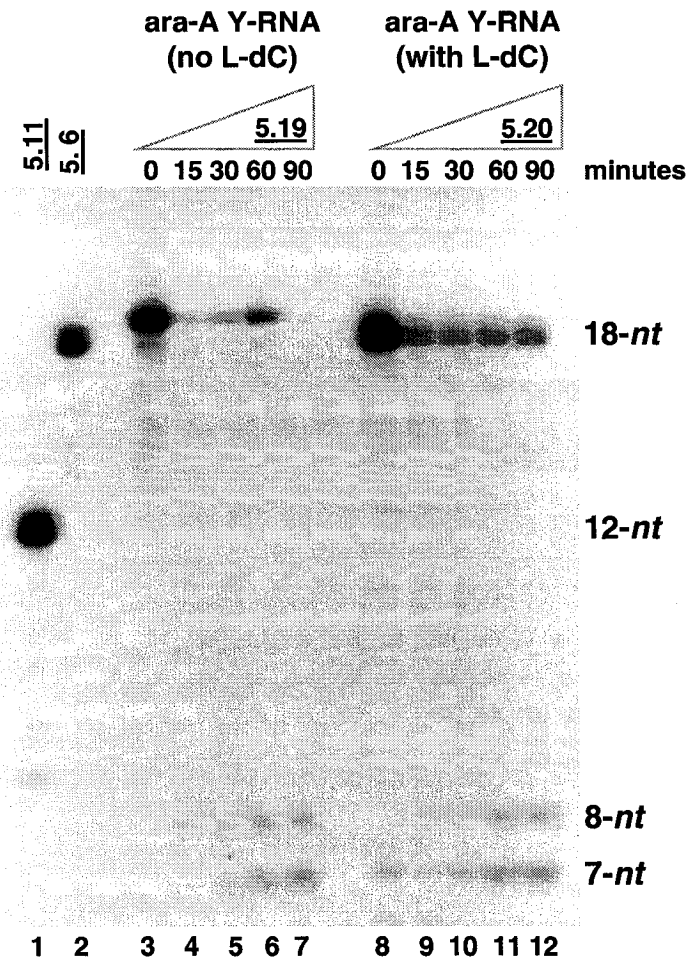


Figure 5.13: Effect of incorporating an arabino-adenosine (aA) branchpoint (**5.19** & **5.20**) and unnatural L-dC termini (**5.20**) on the nuclease stability of Y-RNAs in HeLa nuclear extract. Y-RNAs were incubated with a 25% HeLa extract under splicing conditions. 50 mM KH_2PO_4 was added as a general phosphatase inhibitor. Samples were resolved on a 16% denaturing gel (7M urea) and the bands visualized by autoradiography.

debranching product in the case of **5.20** (**Figure 5.13**), undeniably signifying that the compounds containing an arabinoadenosine branchpoint are not substrates for the 2'-phosphodiesterase HeLa debranching activity. Furthermore, an interesting degradation pattern also emerges on the gels. Initially, as an exonuclease begins its processive degradation of the bNA from the 3'-ends, it likely reaches an interruption site at the 2' and 3'-G-nucleotides preceding the non-natural ara-A, resulting in the production of a 8-nt long sequence (**Figure 5.13**; lanes 3-7). After at least a 30-minute lag time, the exonuclease begins to hydrolyze the phosphodiester bond between the ara-A and the 3'-G

nucleotide and is then unable to process further, thereby accumulating a 7-nt long sequence containing the 2'-phosphodiesterase resistant ara-A/2'-G phosphodiester linkage. The amount of residual bNA **5.20** is probably higher than that demonstrated in **Figures 5.10B** and **5.13** since a considerable amount of 5'-dephosphorylation of the compound was also manifest. As such, bNA **5.20** displayed the highest nuclease resistance of all the bNAs studied (**Figure 5.10B**), predominantly due to the fact that it is insusceptible to hydrolysis by the RNA lariat debranching enzyme.

5.2.3.2. PCR Amplification, Transcription and Splicing of the PIP85.B Substrate Gene

The plasmid containing the PIP85.B splicing substrate gene was obtained from Dr. Andrew MacMillan (University of Alberta). PIP85.B is a viable splicing construct which contains only one adenosine (branchpoint) in the branch region and encodes for the 234-nt sequence shown in **Figure 5.14**.[69] Prior to transcription of the gene, the desired DNA segment was amplified by way of the polymerase chain reaction (PCR) using the thermophilic DNA polymerase, *Taq* polymerase (**Figure 5.15A**).[434] Two DNA primers; M13F (24-nts) and CQ27 (18-nts), were purchased commercially (University of Calgary; Core DNA Services). M13F hybridizes to the region immediately upstream of the 23-nucleotide long T7 RNA polymerase promoter region (position –47), whereas CQ27 is complementary to positions 217-234 (**Figure 5.14**). Amplification was conducted for 30 cycles using an annealing temperature of 45°C for 30 seconds and a 90 second elongation time at 72°C. The amplified dsDNA was purified from the reaction medium by standard phenol/chloroform extraction, and the resultant gene product analyzed on an agarose gel (2%) containing ethidium bromide. Illumination of the gel under UV-light revealed the emergence of an intense, single product band displaying a migratory rate consistent with the expected 281 base-pair length (**Figure 5.15B**). The amplified dsDNA was then transcribed *in vitro* to an active pre-mRNA splicing construct using T7 RNA polymerase (Fermentas). Internal radiolabels were introduced throughout the RNA transcript by way of α-[^{32}P]-UTP as a nucleotide triphosphate source. Analysis of the transcribed RNA by

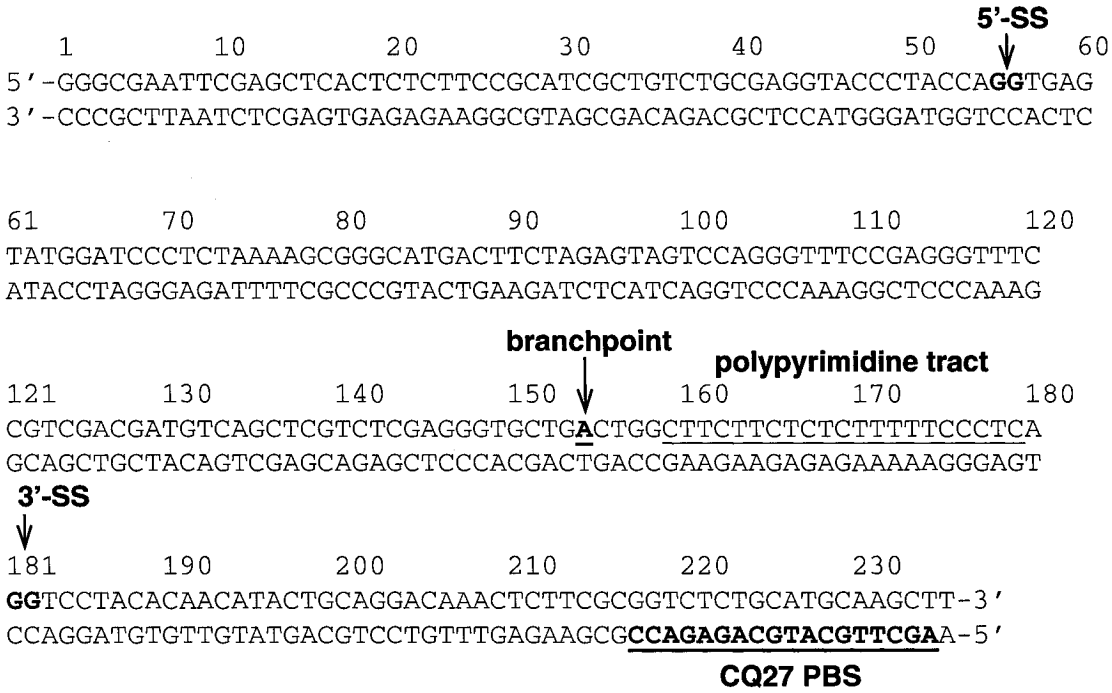


Figure 5.14: Sequence of the PIP85.B splicing substrate gene. The 235-nt sequence is inserted between the T7 promoter region and *Hind* III restriction site of pBS- (Stratagene). When PCR amplified, the T7 promoter region (upstream of +1) and the 234-nt splicing sequence are produced. The 5'- and 3'-splice sites (5'-SS and 3'-SS), branch point adenosine and polypyrimidine tract are shown above. The primer binding sites (PBS) used for PCR amplification are the CQ27 primer (shown) and the universal M13F primer (not shown; located at position -47). CQ27: 5'-AGC TTG CAT GCA GAG ACC-3'; M13F: 5'-CGC CAG GGT TTT CCC AGT CAC GAC-3'.

gel electrophoresis (8%, 7M urea) and autoradiography revealed the inception of a single product species with the expected mobility of the 234-nt RNA (data not shown). Isolation and purification of the transcript by gel extraction and desalting indicated that much of the material had been lost during the purification steps, likely due to the difficulty in handling the low crosslinked gel during the band excision step. Alternatively, purification of the RNA transcript on a commercially available micro-spin column (RNAeasy® Mini Kit; Qiagen) resulted in a significant enhancement of isolated material suitable for use in the pre-mRNA splicing inhibition assays.

Prior to the inhibition assays, the ability for the PIP85.B transcript to splice effectively in HeLa extract was assessed. Pre-mRNA splicing was conducted under the conditions

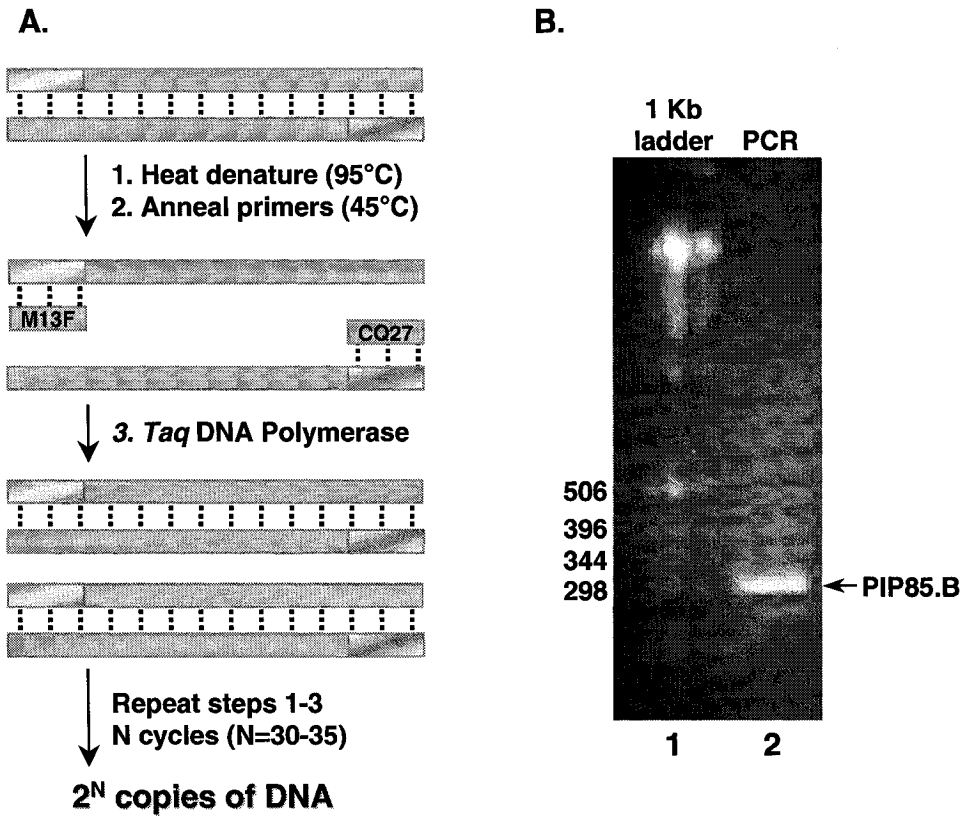


Figure 5.15: PCR amplification of the PIP85.B splicing substrate gene. Panel A: Simplified schematic representation of PCR amplification of a double-stranded DNA substrate using two diverse primers (*e.g.* M13F & CQ27) and the thermophilic *Taq* DNA polymerase. Amplification over a number of number of cycles (N=30-35 cycles) produces an exponential amount of amplified DNA material (2^N). Panel B: Analysis of the PIP85.B PCR product on a 2% agarose gel containing 20 ng/mL of ethidium bromide. The expected length of the PCR product is 281-nt.

described by Grabowski and co-workers[93] and employed 25% v/v HeLa nuclear extract per reaction at a final concentration of 60 mM KCl (see Experimental section 7.12.2.4). The efficiency of transcript splicing was monitored at various time intervals (t=0-90 min). Typically, *in vitro* splicing necessitates a 15-20 min lag-time for spliceosomal assembly and is usually complete after 90 min. As described previously for the yeast system, the splicing course of the substrate gene can be readily perceived by denaturing gel electrophoresis, in which the pre-mRNA substrate, lariat-3'-exon intermediate and the product mRNA and lariat intron are easily distinguished according to their discrete mobilities on the gel (**Figure 5.2**). As seen in **Figure 5.16**, splicing of the 234-nt transcript is indeed active under the employed conditions. Generation of the 109-nt long

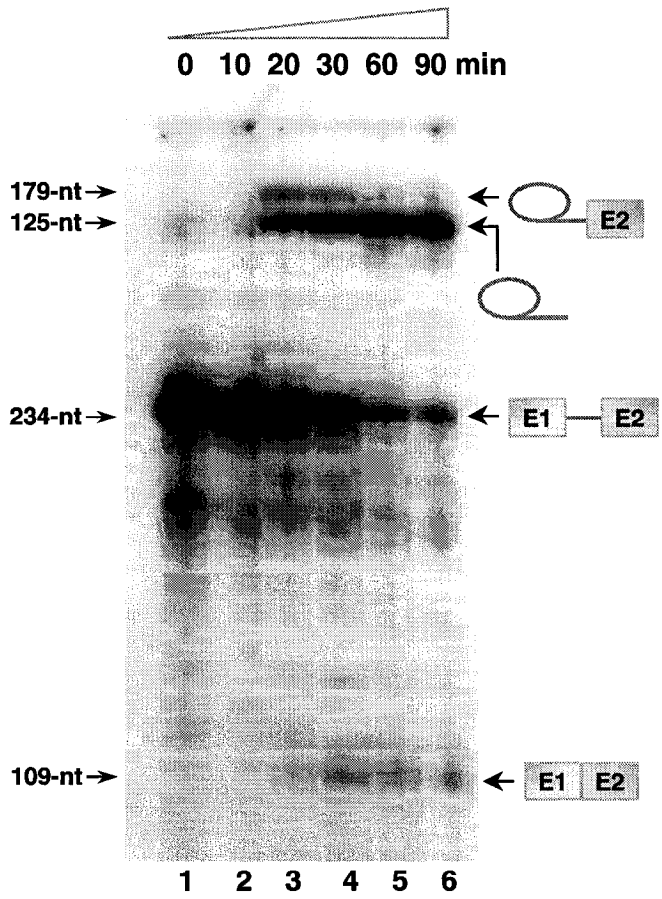


Figure 5.16: *In vitro* pre-mRNA splicing reaction of the PIP85.B substrate gene at various time intervals. Reaction mixtures were partitioned on a 15% (19:1 crosslink) denaturing gel (8 M urea) at constant power (75 Watts). As the reaction time progresses, the amount of mature RNA (mRNA) and lariat intron accumulate, whereas the amount of pre-mRNA and lariat-3'-exon decrease. E1: exon 1; E2: exon 2.

mRNA is seen as early as 20 min into the reaction. Notably, as the reaction time progresses, the amount of pre-mRNA (234-nt) and lariat-3'-exon intermediate (125-nt) suitably decrease, with a concomitant increase in the amount of spliced mRNA product (**Figure 5.16**). This discrete visualization of the lariat intermediate and products will further enable us to dissect which, and if any of the two splicing steps are being affected by the bNA inhibitors. The mobilities of the branched lariat-3'-exon and lariat intron species are indeed anomalous compared to the pre-mRNA, given that their molecular weights are substantially less. The branched topology of these structures likely imposes a frictional interaction with the gel matrix, causing them to boast a retarded electrophoretic displacement.[94] Whereas exogenous bNAs are rapidly debranched by the 2'-phosphodiesterase activity present in HeLa extracts, the lariat intron species (125-nt) accumulates under the splicing conditions described (**Figure 5.16**).[94,100,109] This consistently observed result indicates that the lariat RNA structure is protected from debranching by factor(s) present in the splicing pathway, since previous work in which the lariat-intron has been deproteinized and then added back to a splicing reaction show that it is effectively linearized by cleavage at the 2'-position.[109] Quite possibly, this validates that a spliceosomal factor binds to the branchpoint region and remains bound after intron-lariat excision, thereby protecting the 2',5'-bond from hydrolysis by the lariat debranching activity. This identical factor may be what is recognizing the synthetic bNA substrates *in lieu* of the native lariat during the *in vitro* splicing inhibition assays. Since the transcript was seemingly active under the splicing conditions employed, an investigation of the inhibitory potential of each of the bNAs shown in **Table 5.2** was warranted.

5.2.3.3. Inhibition of *In Vitro* pre-mRNA Splicing by bRNAs in Mammalian (HeLa) Nuclear Extract

Splicing inhibition assays were conducted as described in section 7.12.2.4 of the Experimental. The pre-mRNA transcript was supplemented with 5-20 μM of cold linear, V-shaped or Y-shaped oligonucleotide prior to the addition of splicing cocktail which

contained the HeLa nuclear extract (final concentration: 25% v/v). Only those substrates containing stabilizing L-dC termini were tested in the assay (except for compound **5.19**) since these were shown to be the most stable entities under splicing conditions. Following incubation, the protein content of the splicing reactions was digested and removed by Proteinase K treatment (Fermentas) and phenol/chloroform extraction, prior to being subjected to denaturing PAGE (15%, 8 M urea). The gels were autoradiographed, and in some instances, the amount of inhibition assessed by quantitation of the lariat-3'-exon intermediate band by densitometry at the various concentrations of inhibitor (UN-SCAN-IT Software program; Silk Scientific). The total amount of radioactivity in each lane was difficult to measure owing to the high amount of pre-mRNA band diffusion, thereby preventing the calculation of real IC_{50} values in the following experiments.

5.2.3.4. Branched Oligonucleotide Recognition is Influenced by the Nature of the Furanose Sugar and Nucleotides Preceding the Branchpoint

Splicing of the pre-mRNA substrate in the presence of increasing concentrations of linear (**5.10**), V-shaped (**5.14**) and Y-shaped (**5.16**) DNA oligonucleotides revealed a similar trend as for the yeast system (**Figure 5.17**). Specifically, none of the DNA oligonucleotides, linear or branched, were capable of inhibiting the formation of the mRNA product at the concentration ranges studied (5-20 μM). This was gauged by both qualitative and quantitative analysis, with which no notable decrease in the amount of lariat-3'-exon or lariat intron product was observed with any of the DNA inhibitors (**Figure 5.17**). Possibly, some inhibition might be observed at higher concentrations of oligonucleotide as seen in the yeast system above, although elevated concentrations (*e.g.* 100 μM) were not tested at this time. A converse trend was observed with the RNA inhibitors, specifically the bRNAs. Similar to the yeast system[212], linear RNA **5.12** was not an effective inhibitor of splicing (**Figure 5.18**; lanes 3-5 and **Figure 5.19**). Conversely, the V-RNA **5.15** was a much better inhibitor at higher concentrations (10-20 μM; **Figure 5.18**; lanes 9-11). The most pronounced effect was undoubtedly seen with the native Y-RNAs (**5.8**-**5.9** and **5.17**). Interestingly, the inhibitory activity of bRNA **5.7**

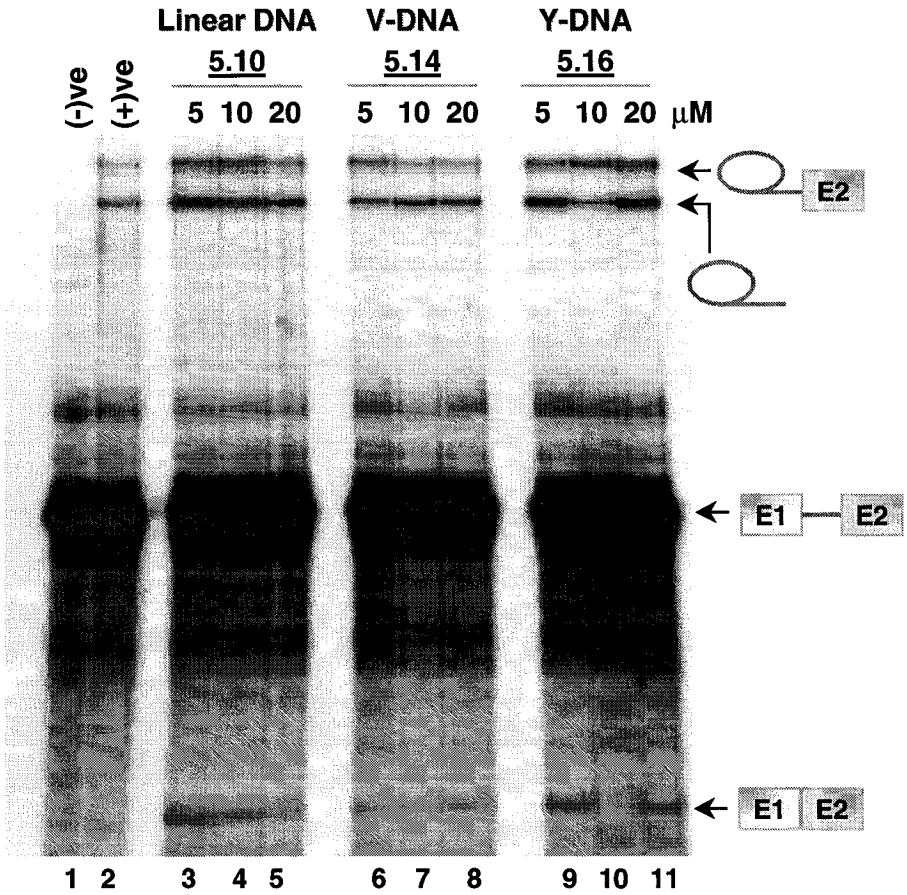


Figure 5.17: Inhibition of pre-mRNA splicing in HeLa nuclear extract with variable concentrations (5-20 μM) of cold linear, V- and Y-shaped DNAs. Splicing reactions were stopped after 30 minutes. Intermediates and products were partitioned on a 15% (19:1 crosslink) denaturing gel (8 M urea) and visualized by autoradiography. The negative control [(-)ve] represents the pre-mRNA alone. The positive control [(+)ve] is the spliced RNA in the absence of any inhibitor.

was seemingly less than that observed for its contingent bNAs, which merely contained increased numbers of stabilizing L-dC inserts (**Figure 5.18**; compare lanes 12-14 with lanes 15-23). Given that the nuclease stability profiles of all the bNAs containing the unnatural termini were very similar, it is safe to conclude that this was just an artifact of the reaction conditions utilized for this particular assay. Nonetheless, a decrease in the amount of lariat-3'-exon was visibly manifest with increasing concentration of all the bRNAs which demonstrated inhibitory potencies (*i.e.* IC_{50}) in the 8-10 μM range

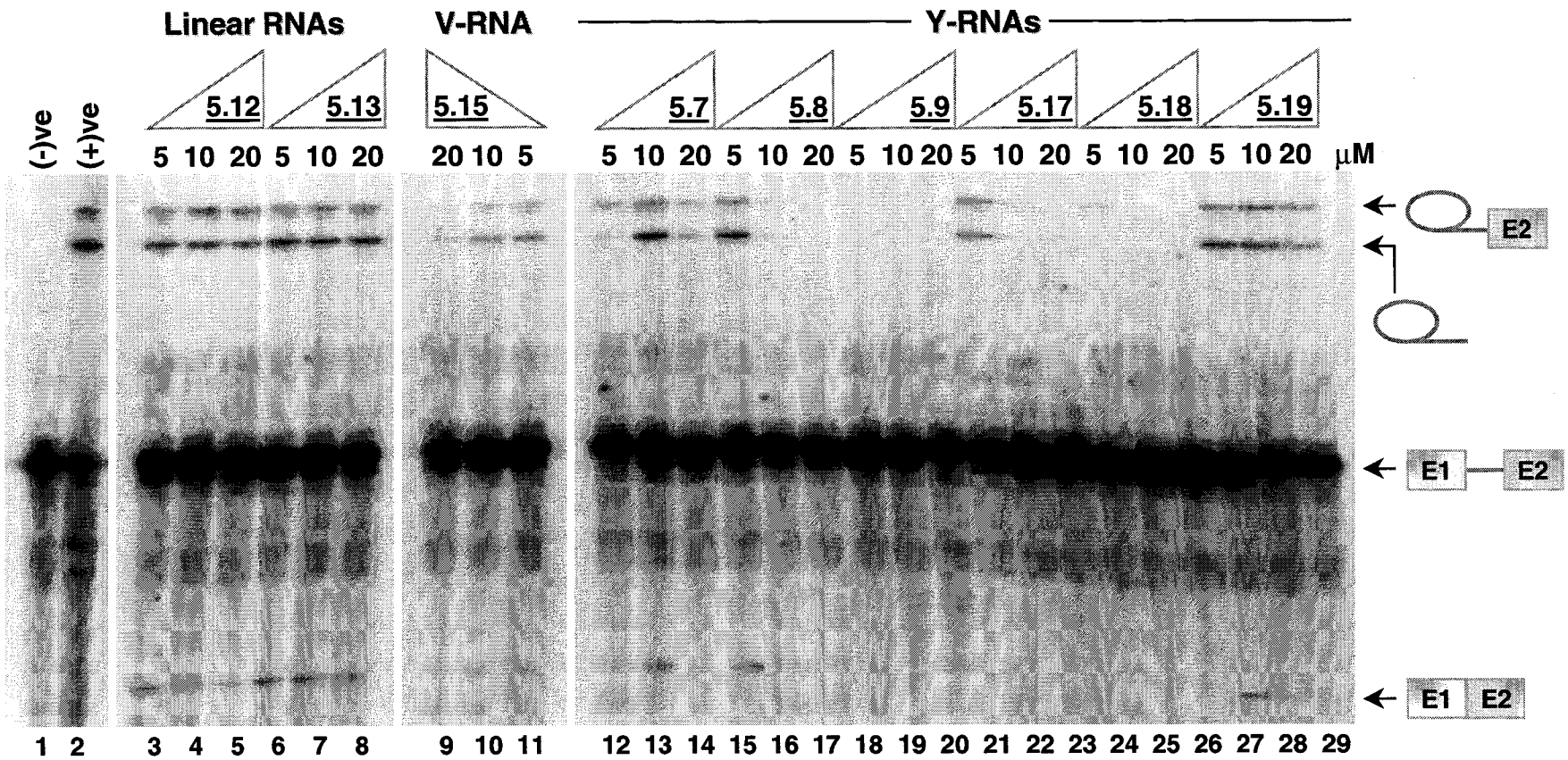


Figure 5.18: Inhibition of pre-mRNA splicing in HeLa nuclear extract with variable concentrations (5-20 μM) of cold linear, V- and Y-shaped RNAs. Splicing reactions were stopped after 30 minutes. Intermediates and products were partitioned on a 15% (19:1 crosslink) denaturing gel (8 M urea) and visualized by autoradiography. The negative control [(-)ve] represents the pre-mRNA alone. The positive control [(+)ve] is the spliced RNA in the absence of any inhibitor.

(**Figures 5.18** and **5.19**). The aforementioned results point to many significant conclusions. Evidently, the lack of inhibition imparted by the branched DNA substrates, especially the Y-DNA **5.16**, indicates once again that branch recognition is not only contingent on the presence of a defined branched core, but also highly dependent on either the 2'-hydroxyl or the predominant C3'-*endo* sugar pucker of the RNA extensions off the branchpoint. Furthermore, seeing that the V-RNA **5.15** did not display as pronounced of an inhibitory potency as the Y-RNA substrates, further establishes the importance of the nucleotides upstream of the branchpoint for adequate binding by the putative branch recognition factor not only in yeast, but in HeLa as well. Given that the lariat-3'-exon and lariat intron species were adequately resolved by gel electrophoresis,

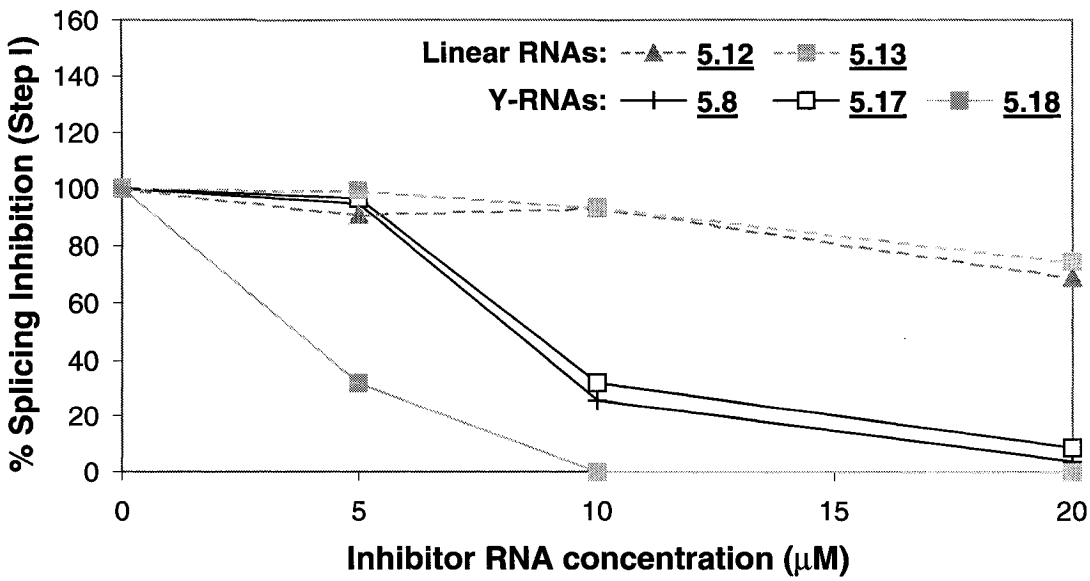


Figure 5.19: Chart demonstrating the inhibition of pre-mRNA splicing in HeLa nuclear extract with variable concentrations of representative linear RNAs (**5.12** and **5.13**) and Y-RNAs (**5.8**, **5.17** and **5.18**). The % splicing inhibition (Step I inhibition) was calculated by densitometric quantitation (UN-SCAN-IT software, Silk Scientific) of the lariat-3'-exon intermediate product band and normalized with respect to the product band in the lane lacking any inhibitor (positive control).

we were capable of dissecting which of the two chemical steps (either Step I or Step II) of the splicing reaction were being affected by the bRNAs. The concurrent decrease in the amount of lariat-3'-exon product (Step I product) and lariat intron (Step II product) undeniably ascertains that binding and sequestering of the branch recognition factor by

the bNAs is an early event, which occurs prior to the first chemical step of splicing. Potentially, an early recognition element such as the p14[108] or mBBP[408,409] protein factors or even the U2 snRNP may be preferentially recognizing the synthetic branched molecules over the native pre-mRNA. The U2 snRNP is a likely candidate for such a recognition event since it is well established that it partakes in early spliceosomal assembly.[66,67] Additionally, it is well recognized that the spliceosomal RNA (U2 snRNA) makes direct contact with the branchpoint, as substantiated by a site-specific crosslinking interaction between a pre-mRNA substrate and the U2 snRNA.[59] Most importantly, the U2 snRNA has also been shown to remain bound to the lariat-3'-exon intermediate following the first transesterification reaction[111], indicating that the folded snRNA structure likely contains a branch recognition domain. As such, it is highly probable that this U2snRNA/lariat-3'-exon interaction is responsible for debranching resistance demonstrated by the lariat intermediate during the splicing reaction.[94,100,109] Whatever the nature of this reputed splicing factor, it is clearly apparent that it demonstrates a much higher affinity for the Y-RNA structures as opposed to linear constructs of similar sequence composition (compare **5.12** and **5.8**). It is therefore highly probable that an indispensable protein or RNA component mixture binds the Y-RNAs preferentially during early assembly, thereby sequestering it from the spliceosomal machinery and inhibiting the first step of the splicing reaction.

The inhibitory activity of the 2'-OMe ribopyrimidine-containing linear (**5.13**) and Y-RNA (**5.18**) were also suitably assessed given their ample nuclease resistance, and the predominant C3'-*endo* (RNA-like) sugar conformation. As mentioned previously, antisense oligonucleotides containing 2'-OMe modifications to the ribose sugar have been effectively exploited for the isolation of ribonucleoprotein complexes, specifically snRNAs.[394,395] Their utility stems from the fact that they are highly nuclease resistant and generally form more stable hybrid duplexes with a complementary RNA sequence than the native RNA.[432,433] Typically the increase in thermal stability of such 2'-OMe constituents is on the order of 1-2°C per modification.[432] The linear 2'-OMe pyrimidine-containing RNA **5.13** demonstrated a nearly identical inactivity towards

splicing inhibition as its all RNA correlative **5.12** under the concentration range studied (**Figure 5.18**; lanes 6-8 and Figure 5.18). Alternatively, the 2'-modified bNA **5.18** displayed significant inhibitory potency (**Figure 5.18**; lanes 24-26 and **Figure 5.19**). Its activity, in terms of step one inhibition, was at least twice as potent as the most active of the bRNAs containing no 2'-modifications, with an IC_{50} of approximately 5 μM (**Figure 5.19**). This result suggests that the C3'-*endo* sugar conformation retained in all the constituents of 2'-modified bNA **5.18** allows for extremely effective recognition by the reputed branch recognition factor. Additionally, should this factor be an snRNA (*i.e.* U2 snRNA), the enhancement in activity may be due to the fact that tighter binding in the complex is achieved owing to the 2'-OMe substituents.[432]

5.2.3.5. Inhibitory Activity is Maintained in a Branched RNA Containing a Modified Arabino-Adenosine Branchpoint

Although remarkable inhibitory potencies were achieved with the bRNAs described in the previous section, nuclease degradation profiles (**Figure 5.10-5.12**) undeniably established that 2'-debranching of the substrate bNAs was a serious drawback. This hydrolytic activity essentially reduces the effective concentration of bNA inhibitors during the splicing assays. As such, the concentrations reported are probably much lower than those observed in reality. Given this limitation, the debranching resistant bNAs **5.19** and **5.20** were also investigated for their inhibitory potential (**Figure 5.20**). Under the splicing inhibition conditions described previously, bNA **5.19**, which did not contain a 3'-L-dC stabilizing insert, appeared to inhibit step one of the RNA splicing reaction at 20 μM concentration (**Figure 5.20**; lanes 3-5). Conversely, a much lower concentration of bNA **5.20** was required to afford complete splicing abolition under the conditions employed (**Figure 5.19**; lanes 6-11). Furthermore, quantitative analysis of the lariat-3'-exon intermediate indicated that step one inhibition was occurring with as little at 0.5 μM of the bNA **5.20**. This generous inhibitory activity seemingly indicates that the non-native arabinoadenosine has little bearing on recognition by the branch recognition factor. In comparison to **5.18**, the best inhibitor from the ribose-containing branchpoint series, **5.20**

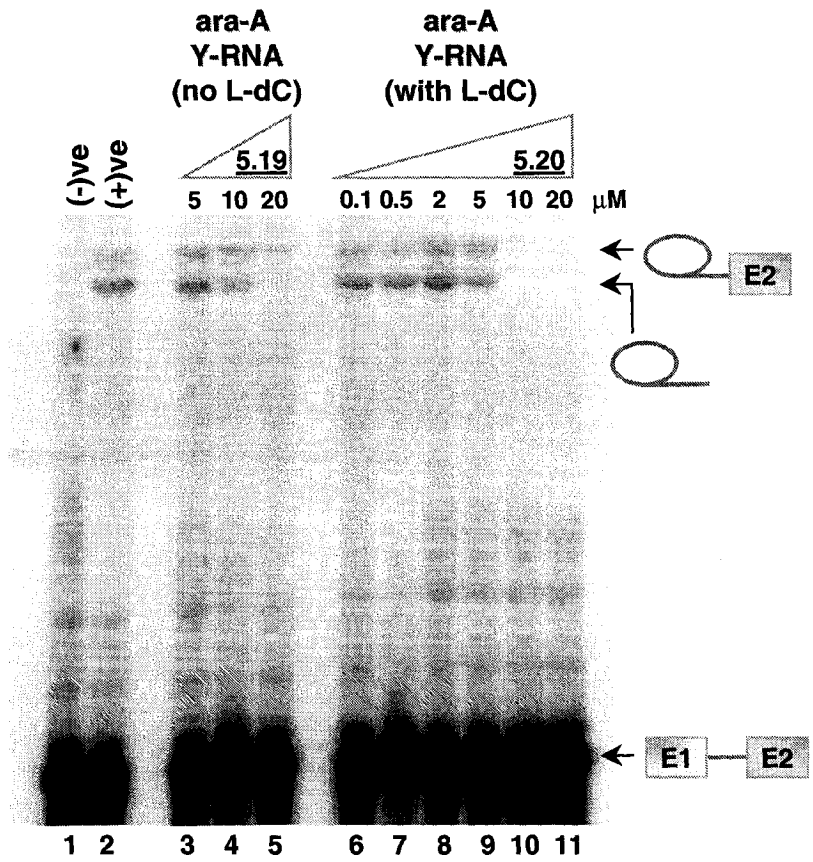


Figure 5.20: Inhibition of pre-mRNA splicing in HeLa nuclear extract with variable concentrations of cold Y-RNAs containing an arabino-adenosine branchpoint (**5.19** and **5.20**). Splicing reactions were stopped after 30 minutes. Intermediates and products were partitioned on a 15% (19:1 crosslink) denaturing gel (8 M urea) and visualized by autoradiography. The negative control [(-)ve] represents the pre-mRNA alone. The positive control [(+)ve] is the spliced RNA in the absence of any inhibitor.

appeared to be slightly less potent at identical concentrations (*i.e.* 5 μM). While the sugar conformation adopted by the arabinose sugar at this branch position has not been validated, studies on a branched trinucleotide diphosphate containing an arabinoadenosine branchpoint have implied the adenine heterocycle is capable of a base-base stacking interaction with the residue at the 2'-position, similar to the native ribose branchpoint.[314] Mutation of a conserved riboadenosine unit for an arabinoadenosine moiety in a pre-branched pre-mRNA substrate has formerly exposed compelling evidence that this unit is capable of occupying a "bulged" position, similar to the riboadenosine residue when

bound to the GUAGUA region of the U2 snRNA.[69] These combined results suggests that a similar point of local A-like distortion is likely observed with the arabinose modification, and may also act as a underlying motif for branch recognition. Introduction of an apurinic branchpoint, which is incapable of any stacking, may provide further support that the 2'-base stacking interaction is an essential element for bNA recognition during splicing.

5.2.3.6. Small, Branched Oligonucleotides are Poor Inhibitors of RNA Splicing

Additional evidence supporting that branch recognition requires the presence of a fully defined branchpoint including 5'-, 2'- and 3'-extensions was obtained with small, branched oligonucleotides, specifically trimers and tetramers (**Table 5.3**). Trimers and tetramers were provided by Dr. Masad Damha and purified by Dr. Katya Viazovkina of our research group. Incubation of variable concentrations (5-20 μM) of the assorted

Table 5.3: List of linear, V- and Y-shaped small oligonucleotides (*i.e.* trimers, tetramers and decamer) utilized as inhibitors of *in vitro* pre-mRNA splicing in mammalian nuclear extract (HeLa)

Code	Sequence (5'→3')	Topology
-	AUC[a]	Linear RNA Trimer
-	$A^{2',5'}(U)_{3',5'}U^a$	V-RNA Trimer
-	$A^{2',5'}(U)_{3',5'}G^a$	V-RNA Trimer
-	$A^{2',5'}(G)_{3',5'}U^a$	V-RNA Trimer
-	$UA^{2',5'}(U)_{3',5'}U^a$	Y-RNA Tetramer
5.21	$AA^{2',5'}(G)_{3',5'}G$	Y-RNA Tetramer
5.22	$cUAA^{2',5'}(GUc)_{3',5'}GUc$	Y-RNA Decamer

Notation: large cap letters=ribonucleotide residues; *c*=unnatural L-2'-deoxycytidine (L-dC); $^{2',5'}$X=2',5'-phosphodiester linkage; $_{3',5'}$X=3',5'-phosphodiester linkage. [a]These sequences were provided by Dr. Masad Damha and deprotected and purified by Dr. Katya Viazovkina.

compounds under pre-mRNA splicing conditions demonstrated that in almost all cases, inhibitory activity was eradicated. As expected, no activity was observed with the linear trimer AUC (**Figure 5.21**; lanes 3-5). When the trimers featured a branch core, as in the case of A(U)U, A(U)G and A(G)U, no inhibitory potential was also established under the concentration range investigated (**Figure 5.20**; lanes 6-14). Although the A(G)U trimer

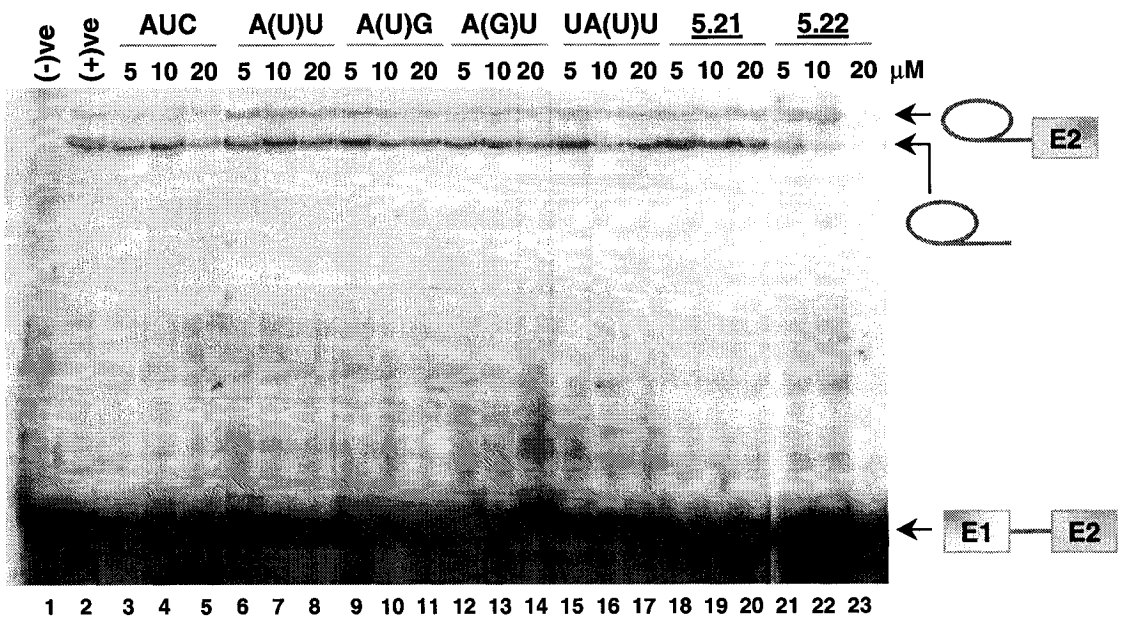


Figure 5.21: Inhibition of pre-mRNA splicing in HeLa nuclear extract with small branched oligonucleotides. Cold linear, V-shaped and Y-shaped RNAs were incubated with the pre-mRNA transcript under splicing conditions. Splicing reactions were stopped after 30 minutes. Intermediates and products were partitioned on a 15% (19:1 crosslink) denaturing gel (8 M urea) and visualized by autoradiography. The negative control [(-)ve] represents the pre-mRNA alone. The positive control [(+)ve] is the spliced RNA in the absence of any inhibitor. Bracketed residues indicates nucleotides linked by a 2',5'-phosphodiester bond.

contained the conserved 2'-G unit found in the natural lariats, it was still unable to promote inhibition of splicing activity in the HeLa surroundings. The same was true for the branched tetramer **5.21**, which possessed the 2'-G residue in the structure (**Figure 5.21**; lanes 19-21). In fact, neither of the two branched tetramers (*i.e.* UA(U)U and **5.21**) were active splicing inhibitors, regardless of the fact that they contained a single 5'-nucleotide off the branch. Contrarily, the longer decanucleotide branched species **5.22** demonstrated some potency at higher concentrations (*i.e.* 10-20 μM). This further

supports the notion that branch recognition is dependent upon the sequences directly adjacent to the branchpoint nucleotides. As with the V-RNA (**5.15**), liberal inhibition required that the branched compounds bear a 5'-extension. Whether the 2'- and/or 3'-appendages play a significant role in terms of factor recognition remains to be assessed.

5.3. CONCLUSIONS

The comprehensive results described in the preceding sections clearly point towards the utility of suitably stabilized branched oligonucleotides as agents for the study of branchpoint recognition during pre-mRNA splicing. Most notably, those substrates containing a minimum of one unnatural L-dC unit at their 3'-termini were highly stabilized against exonuclease hydrolysis in both the yeast and HeLa extract milieus. Our findings also corroborate previous evidence[208,212] that fully-formed branched oligonucleotides, specifically bRNAs are potent inhibitors of the splicing pathway. The requirement of a fully formed branch, especially the need for a 5'-extension, was supported by the inactive inhibitory profiles demonstrated by V-shaped structures and small, branched trimers and tetramers. Additionally, a bRNA containing 2'-OMe ribopyrimidine insertions (**5.18**), was also an exceptionally potent inhibitor. Apart from the mildly increased stabilization exuded by this compound, presumably against endonucleases present in the HeLa extract, the compound also displayed more potent splicing inhibition compared to the native bRNAs which lacked the 2'-OMe modification. Seemingly, this enhancement may be a result of tighter binding to the branch recognition factor, specifically if it is an snRNA. Our results with the arabinose-containing branchpoint (**5.20**) affirm that this type of branched oligonucleotide is not a substrate for the 2'-scissle activity (debranching activity) found in HeLa nuclear extract.[213] Moreover, this trait appears to be beneficial since it would substantially increase the effective concentration of inhibitor present throughout the reaction. The potent activity of this modified bRNA as well as preceding evidence on arabinose containing bNAs and pre-mRNAs[69,314] indicates that the branchpoint likely assumes a similar conformation to the native RNA branchpoint, thereby serving as a local distortion point for recognition

by the putative branch recognition factor. Furthermore, the decrease in the amount of lariat-3'-exon intermediate points towards a mechanism of step one inhibition during the splicing reaction. As such, it is speculated that the branch recognition element is an early binding factor, which recognizes the bNAs during the initial spliceosomal assembly events. The U2 snRNP factor is known to be an early recognition element, and its interaction with the branchpoint after lariat formation (after step one) has been suitably documented.[111] Given that linear RNA oligonucleotide constructs which bear the same brancpoint consensus sequence as the Y-shaped structures, were not adequate inhibitors of splicing, it seems feasible that branch recognition might not only rely on the U2 snRNP, but quite possibly a U2 snRNP associated factor which plays a more pivotal role in recognition of the lariat species.

Given this abundance of information, studies towards the biochemical analysis of the branch recognition factor are now possible. In essence, a variety of techniques can be exploited to determine the nature of this spliceosomal factor. A particularly useful tool would be the site-specific introduction of a photoreactive 4-thio-uridine unit into a bRNA, to determine if any explicit crosslinking interactions are possible with the regions surrounding the branchpoint. In addition, such bNAs can potentially be employed in the affinity selection of this unidentified factor by biotinylation of the 5'-terminus and hydridization to a steptavidin-immobilized matrix. The nature of this factor as a protein, RNA or possibly both can then be determined by standard biochemical methods such as Western blotting or Northern hybridization.

5.4. MODULATION OF BCL-X ALTERNATIVE SPLICING BY LINEAR AND BRANCHED ANTISENSE OLIGONUCLEOTIDES

5.4.1. Alternative Splicing Regulation of Programmed Cell Death

Alternative splicing is an elaborate cellular process devised by higher-order eukaryotic genomes, and is considered to be the primary source of human proteomic diversity.[124,125] Specifically, alternative splicing allows for the selection of different combinations of 5'- and 3'-splice sites within a pre-mRNA transcript, consequently allowing for the expression of multiple mRNAs encoding distinct protein products with diverse and sometimes antagonistic functions. It is estimated that up to 60% of human genes generate multiple mRNAs by alternative splicing, thereby increasing the coding capacity of a single pre-mRNA.[123]As such, alternatively spliced genes control the production of numerous proteins displaying diverse biological effects, with one form possibly contributing to a human disease and another form preventing it. One such set of alternatively spliced protein products constitute the Bcl-2 family of proteins. Bcl-2 and its related protein, Bcl-x are directly linked to regulation of cell death.[136,435] Programmed cell death or apoptosis is the orderly and elegant signaling process by which cells commit suicide, and is an essential component of organismal development. Apoptosis controls cellular differentiation, cell numbers, and the removal of aged, damaged or autoimmune cells. A complex choreography of death signals interact with one another and interface with multiple signaling pathways, creating a cascade of proteins involved in cell death (**Figure 5.22**).[436,437] In essence, this strictly controlled program of cell suicide has been conserved throughout evolution, and is responsible for the prevention of uncontrolled cell proliferation. In contrast to necrotic cell death, which involves a local inflammatory response, apoptosis is defined by a reduction in cell volume and membrane blebbing, resulting in shriveled cell bodies that are eliminated by nearby phagocytes.

Numerous genes have been described that function to regulate susceptibility to apoptosis. Bcl-2, the founding member of this family of proteins is involved in the mitochondrial

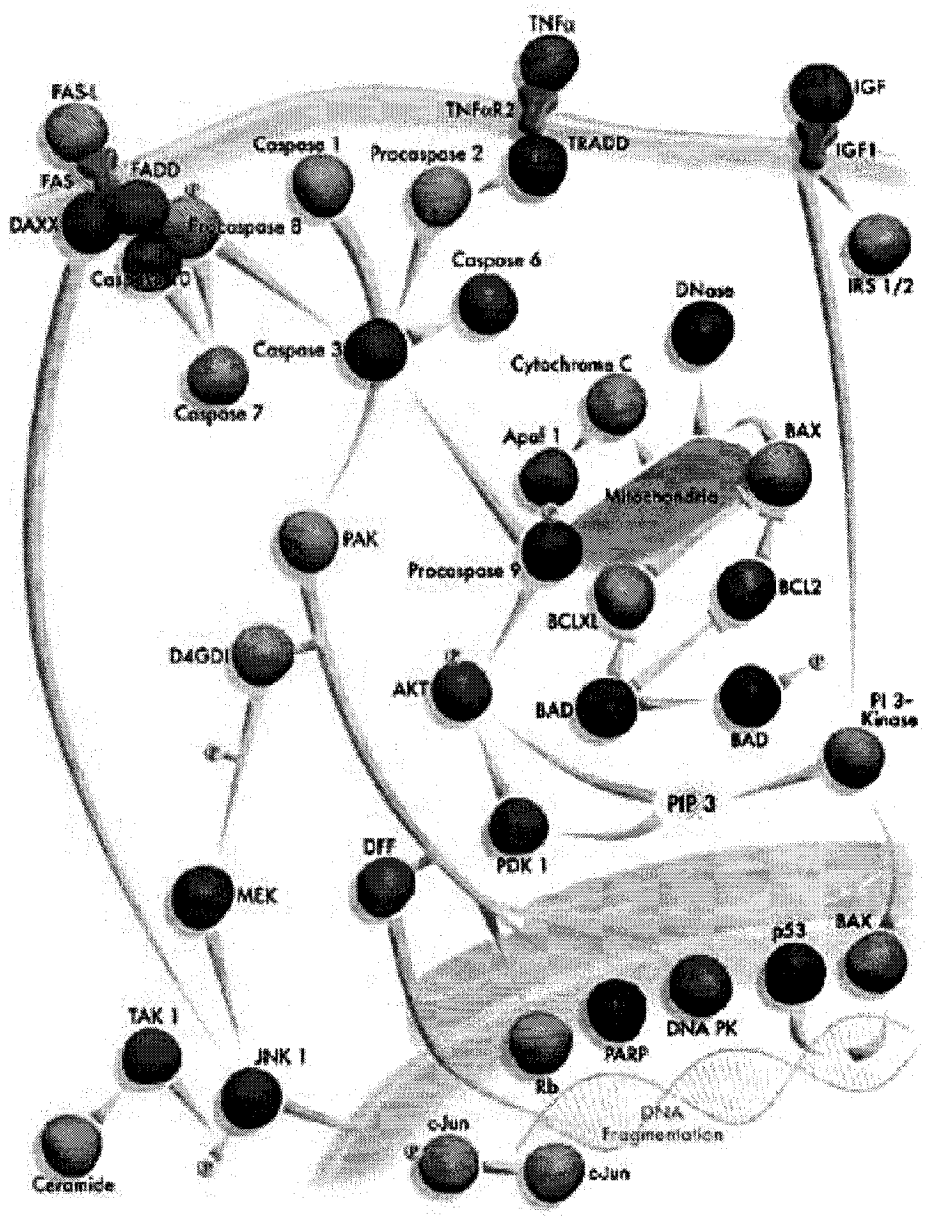


Figure 5.22: Complex interacting signaling events surrounding cellular apoptosis. Adapted from: Lewis, R. Deciphering Death's Circuitry. *The Scientist.* **2003** (March 24th issue), pp. 32-33.

apoptotic pathway.[436] Bcl-2 was initially identified by its overexpression in a form of human lymphocyte cancer.[438] This oncogene was quite novel since rather than stimulating proliferation of a cancer cell, it instead enhanced its ability to survive under suboptimal conditions (*i.e.* anti-apoptotic protein). Approximately eight years later, a similar gene, *bcl-x*, a member of the Bcl-2 family was also exposed [136]. The presence of

two alternative 5'-splice sites in the second exon of the *bcl-x* gene results in two alternatively spliced variants, which encodes two distinct protein isoforms, namely Bcl-x_L (long form) and Bcl-x_S (short form) (**Figure 5.23**). As such, Bcl-x_S lacks an internal 63 amino acid residues that encompass the region of greatest sequence homology between Bc-x_L and Bcl-2. Contrary to Bcl-x_L which has been shown to hinder apoptosis *via* heterodimerization with various pro-apoptotic proteins such as BAX and BAK[439,440], Bcl-x_S is itself pro-apoptotic[136] and functions by counteracting cell survival proteins such as Bcl-x_L and Bcl-2.[138,441]. Whereas Bcl-x_L is predominantly expressed in many human carcinomas and lymphomas[442] its spliced variant Bcl-x_S is down-regulated in transformed cells, and is up-regulated in cells undergoing apoptosis. Therefore, the *bcl-x* gene is a clear example of how alternative RNA splicing can result in the generation of proteins with antagonistic cellular functions.

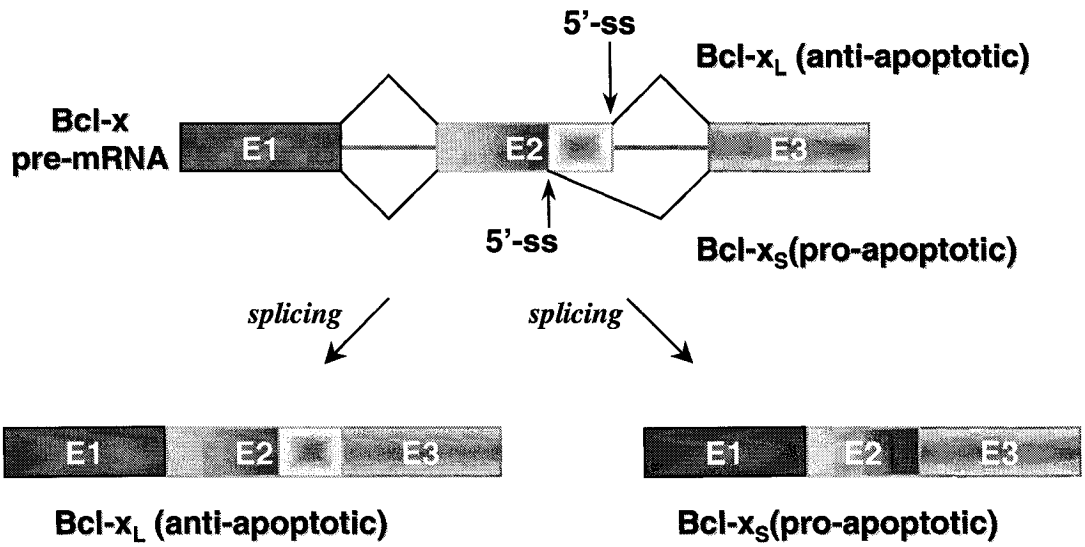


Figure 5.23: Alternative splicing of the Bcl-x pre-mRNA. Translation of Bcl-x_S results in an protein that is 63 amino acids shorter than that encoded by Bcl-x_L. The alternative 5'-splice sites (5'-ss) are denoted for both the Bcl-x_L and Bcl-x_S isoforms.

Since overexpression of Bcl-x_L in cancerous cells is associated with decreased apoptosis in malignant tumors under appropriate stimuli, resistance to chemotherapy, and a lower incidence of remission, an appropriate drug intervention program that decreases the amount of Bcl-x_L and concomitantly increases the amount of expressed Bcl-x_S levels,

could prove to be an effective anticancer therapeutic. A valuable intervention method towards the pharmacological regulation of splice site selection is antisense technology.[142,143,443,444] In the antisense method, a synthetic oligonucleotide is designed to be complementary to the sense portion of an RNA, thereby hindering expression of an unwanted gene and possibly promoting expression of a functional or beneficial gene. Antisense[397,398] is the only drug discovery technology capable of targeting a gene's RNA in a sequence-specific fashion, thereby regulating alternative splicing.[399] Whereas this technology has primarily been exploited for the down-regulation of gene expression by either blocking translation (*i.e.* translation arrest) or promoting RNase H-mediated degradation of viral, oncogenic or other targeted RNA messages, its applicability towards the redirection of splice site selection is also noteworthy. Specifically, such constructs were first shown to be effective in redirecting the splicing of an aberrant splice site in the thalessemic β-globin gene thereby upregulating production of the normal protein, hence restoring gene function.[400,401] In most cases, studies are still at the pre-clinical stage, however a few noted antisense compounds have found their way to clinical trials for treatment against a variety of disorders.[445,446] In 2002, an international phase III clinical trial initiated with an 18-nucleotide phosphorothioate antisense compound developed by Genta and Aventis Pharma (G3139), targeted toward anti-apoptotic Bcl-2 for the treatment of malignant melanoma.[447] It is speculated that the mechanism of action of this compound is to downregulate Bcl-2 expression, thereby promoting apoptosis and sensitizing cells to conventional chemotherapeutic treatments. Recently, a heavily cited report emerged concerning the use of antisense oligonucleotides, specifically phosphorothioates containing uniform 2'-O-(2-methoxy)ethyl (2'-MOE) modifications to redirect splicing of the Bcl-x pre-mRNA from the Bcl-x_L 5'-splice site to that of Bcl-x_S (**Figure 5.23**).[141] The most effective construct, which hybridized 15 bases upstream of the Bcl-x_L 5'-splice site, was capable of increasing the expression levels of Bcl-x_S while downregulating Bcl-x_L expression in a dose- and time-dependent fashion. This study and others have inferred that such modified oligonucleotides likely inhibit pre-mRNA splicing by obstructing spliceosomal assembly at targeted splice sites forcing it to reposition at an alternate site.

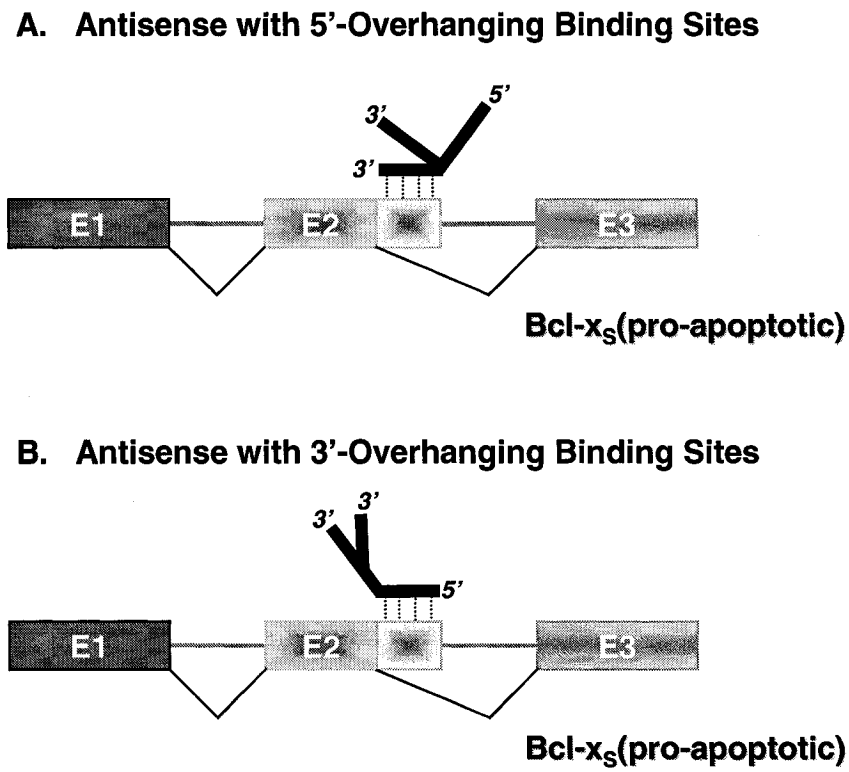


Figure 5.24: Modulation of Bcl-x splicing using antisense oligonucleotides containing linear and branched 5'-overhanging protein binding sites (Panel A) or 3'-overhanging protein binding sites (Panel B). Constructs were designed to effectively downregulate Bcl-x_L splicing, and upregulate the splicing of Bcl-x_S. The anti-Bcl-x regions of the antisense constructs were complementary to the -4 to -23 positions upstream from the 5'-splice site of Bcl-x_L.

While antisense oligonucleotide treatment alone was insufficient for the induction of apoptosis in the cancerous lung epithelial cell line studied, it did however sensitize the cells to both UV-B radiation and cis-platin, an apoptosis-inducing chemotherapeutic.[141] The results suggested that the Bcl-x_S expression levels achieved were insufficient to overcome the anti-apoptotic stimuli in this cell line, and that attainment of higher expression levels in other cell lines could reveal its full pro-apoptotic potential. Indeed, Mercatante and co-workers later demonstrated that antisense 2'-O-methyl phosphorothioate oligonucleotides targeted to a similar region of the Bcl-x pre-mRNA in prostate cancer cells resulted in a marked increase in Bcl-x_S expression and induced apoptosis in the absence of any cytotoxic stimuli.[140] Given that recurrent prostate cancer

tissue samples express higher levels of Bcl-x_L than benign prostate tissue suggests a remarkable specificity for such oligonucleotide constructs as non-cancerous cells bearing low levels of Bcl-x_L should be resistant to treatment.[448] Although antisense oligonucleotides have been successful at redirecting the splicing of Bcl-x, maximal expression levels of 50-60% Bcl-x_S were not obtained in all cell lines.[140,448] As such, we wished to investigate an alternate antisense approach to modulating and interfering with Bcl-x 5'-splice site selection, hopefully promoting more efficient Bcl-x_S formation. This approach uses oligonucleotide constructs containing overhanging binding sites for proteins or ribonucleoprotein complexes either upstream or downstream of the 5'-splice site in exon 2 of Bcl-x (**Figure 5.24**). We anticipated that such antisense constructs would be able to modulate Bcl-x splicing *via* a dual mode of action whereby one segment is complementary to the region encompassing the Bcl-x 5'-splice site, and the non-complementary linear or branched oligonucleotide portion would contain factor binding sequences which could potentially interfere with spliceosomal or non-spliceosomal factors, thus potentially allowing for more pronounced Bcl-x_S expression profiles.

5.4.2. Synthesis of Linear and Branched Factor-Binding Antisense Oligonucleotide Constructs

Linear antisense RNA oligonucleotide constructs (**5.23**-**5.24**, **5.26**-**5.27** and **5.29**-**5.30**; **Table 5.4**) were synthesized on an ABI 381A DNA synthesizer using standard silyl-phosphoramidite chemistry, coupling reagents and conditions (section 7.3). Conversely, branched RNAs (**5.25**, **5.28** and **5.31**-**5.34**; **Table 5.4**) were assembled using our well-established convergent solid-phase methodology as described in the preceding section (section 5.2).[206,233] The conserved adenosine branchpoint was introduced *via* the adenosine bis-phosphoramidite **1.1** under dilute conditions (0.03 M).[204,205] In order to afford maximal branching, high-loading CPG was utilized.[204,233] Additionally, antisense molecules less that 40 nucleotides in length were constructed on a 500 Å CPG whereas longer oligonucleotides (> 40-nt) were assembled on a 1000 Å CPG pore size in order to prevent steric clashing between the growing oligonucleotide strands. As

Table 5.4: List of antisense and non-antisense sequences utilized in the modulation of Bcl-x alternative splicing.

CODE	Sequence (5'→3')	#nts	RNA Topology
Control RNAs			
5.23	GCC GCC GUU CUC CUG GAU CC	20	Antisense Control
5.24	AUA GGC ACU GA GUU GGU AUG*a*	21	Non-Antisense
5.25	AUA GGC ACU GA$^{2'5'}$(GUU GGU AUG*a*)$_{3'5'}$GUU GGU AUG*a*	31	Non-Antisense
Antisense with 5'-Overhangs			
5.26	AAU GUC UGC UAC UGG AAG AAU GCC GCC GUU CUC CUG GAU C*c*	41	Linear
5.27	UGG GUU UCU GAU AGG CAC UGA GCC GCC GUU CUC CUG GAU C*c*	41	Linear
5.28	UGG GUU UCU GAU AGG CAC UGA$^{2'5'}$(GCC GCC GUU CUC CUG GAU C*c*)$_{3'5'}$GCC GCC GUU CUC CUG GAU C*c*	61	Y-RNA
Antisense with 3'-Overhangs			
5.29	GCC GCC GUU CUC CUG GAU CC AAU GUC UGC UAC UGG AAG AAU*c*	42	Linear
5.30	GCC GCC GUU CUC CUG GAU CC AUA GGC ACU GA GUU GGU AUG*a*	41	Linear
5.31	GCC GCC GUU CUC CUG GAU CC AAU GUC UGC UA$^{2'5'}$(AAU GUC UGC U*c*)$_{3'5'}$AAU GUC UGC U*c*	53	Y-RNA
5.32	GCC GCC GUU CUC CUG GAU CC AAU GUC UGC UA$^{2'5'}$(GUU GGU AUG A*c*)$_{3'5'}$GUU GGU AUG A*c*	53	Y-RNA
5.33	GCC GCC GUU CUC CUG GAU CC AUA GGC ACU GA$^{2'5'}$(AAU GUC UGC UAC UGG AAG AAU*c*)$_{3'5'}$AAU GUC UGC UAC UGG AAG AAU*c*	75	Y-RNA
5.34	GCC GCC GUU CUC CUG GAU CC AUA GGC ACU GA$^{2'5'}$(GUU GGU AUG A*a*)$_{3'5'}$GUU GGU AUG A*a*	53	Y-RNA

Notation: Underlined sequences are complementary to the 5'-splice site (5'-ss) of Bcl-x_L (positions –4 to –23 upstream of the 5'-ss). The overhang extensions either protrude over the second intronic segment (5'-overhangs) or the second exonic portion (3'-overhangs) of the Bcl-x *si* mini gene (**Figure 5.24**). Capital letters represent ribonucleotide units. $^{2',5'}$X= a 2',5'-phosphodiester bond; $_{3',5'}$X= a 3',5'-phosphodiester bond. *c* and *a* represent unnatural L-dC and L-dA nucleotides respectively.

mentioned previously, a larger pore diameter insures adequate distance between the strands, since smaller pores tend to fill up with the growing oligonucleotide thereby severely reducing the efficiency of coupling further phosphoramidites. Furthermore, 500 Å CPG typically supports a higher loading and density of oligonucleotide chains on the

surface, thereby increasing the steric crowding between neighboring strands. The antisense oligonucleotides were designed to contain a 20-nucleotide portion, which was complementary to the region directly upstream (positions -4 to –23) of the Bcl-x_L 5'-splice site in exon 2 of Bcl-x. Previous work emerging from the laboratory of Dr. Benoit Chabot has shown that targeting the region immediately upstream of the 5'-splice site is more effective at reprogramming alternative splicing that directly targeting the 5'-splice site itself.[449] Antisense molecules contained homogenous ribonucleotide units since RNA-RNA duplexes exhibit high thermal stability and are completely resistant to the endogenous RNase H-mediated destruction of the targeted message.[433] Moreover, the molecules were also devised to contain non-hybridizing portions (*i.e.* 3'- or 5'-overhangs) with specific spliceosomal factor recognition sequences (**5.27**-**5.28**, **5.30** and **5.32**-**5.34**) (**Figure 5.26**). Such sequences included an intronic 5'-splice site region or branchsite sequence of the β-globin pre-mRNA. Control sequences were also formulated to determine the importance of the anti-Bcl-x sequence (**5.24** and **5.25**) and protein binding sequences (**5.26**, **5.29** and **5.31**) on the specificity of antisense modulation.

In an effort to stabilize the molecules against pervasive exonucleases present in HeLa extracts, unnatural L-deoxynucleotide inserts (L-dC and L-dA) were introduced at the 3'-termini of the molecules using the appropriate nucleoside loaded CPG. As seen in the preceding section, one insert of an L-deoxynucleotide unit at the 3'-terminus of an otherwise unmodified synthetic oligonucleotide imparts sufficient resistance against exonuclease hydrolysis. Oligonucleotides were deprotected under standard conditions (section 7.4) along with an ensuing treatment with the desilylation reagent, TREAT-HF to remove the 2'-TBDMS protecting group. Furthermore, the molecules were analyzed and purified by denaturing PAGE (12%, 7M urea), desalted by SEC (section 7.5.5.1) and their nucleotide composition confirmed by MALDI-TOF-MS (section 7.5.6; **Table 7.6**). Analysis of the oligonucleotide products by denaturing PAGE (**Figure 5.25**) revealed that all of the compounds were synthesized with high efficiency, despite the fact that some of the constructs were the longest and largest bRNAs ever realized (*i.e.* **5.28** and **5.31**-**5.33**). Their migratory behaviors on the gel were consistent to those of commercial standards (40 and 50-mers; **Figure 5.25**).

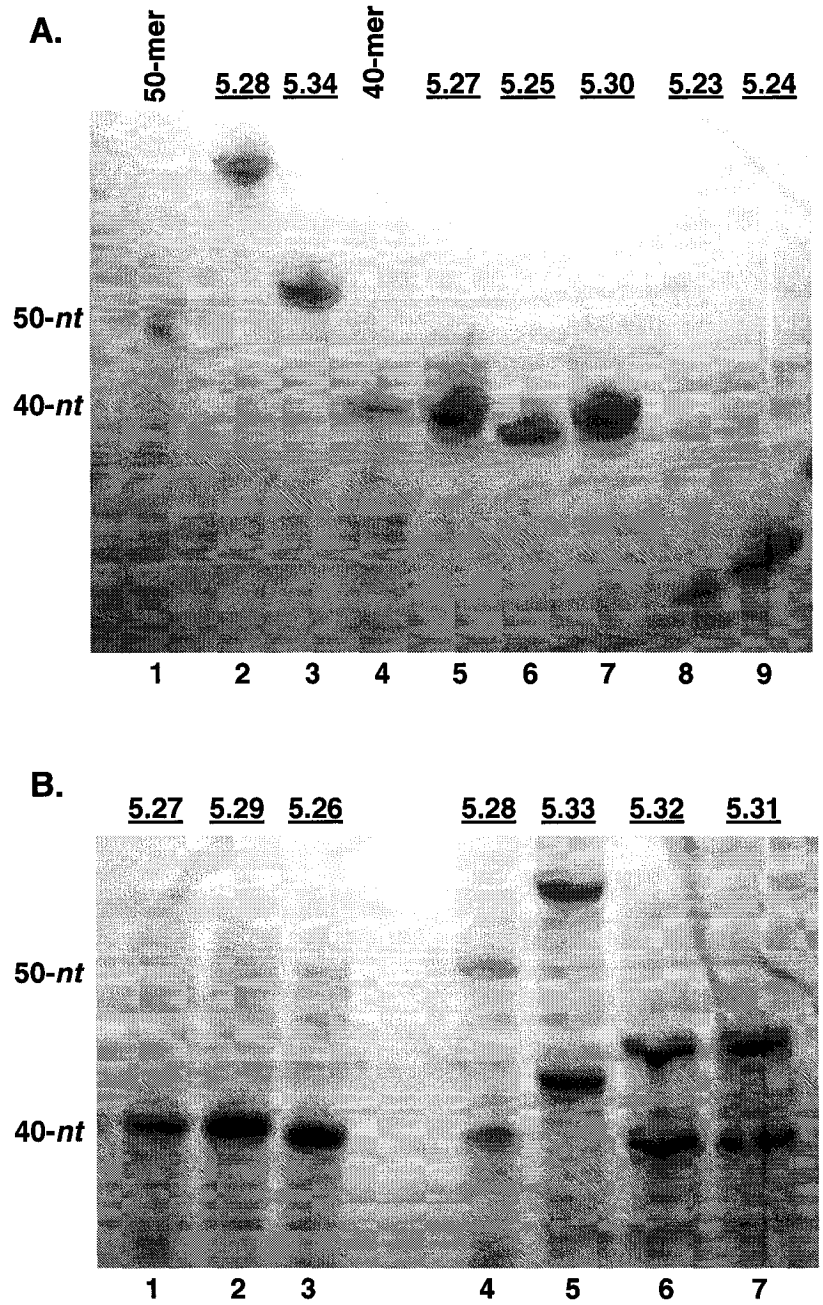


Figure 5.25: Analytical PAGE analysis (12%, 7M urea) of the various linear and branched antisense oligonucleotides utilized in the study containing either 3'- or 5'-overhanging binding sites. Panel A: analytical gel of purified linear and branched RNA samples; Panel B: analytical gel of crude linear and RNA samples. The slower migrating band in the crude branched samples is the linear isomeric failure sequence formed by unsuccessful branching of the bis-phosphoramidite **1.1** to two adjacent support-bound strands.

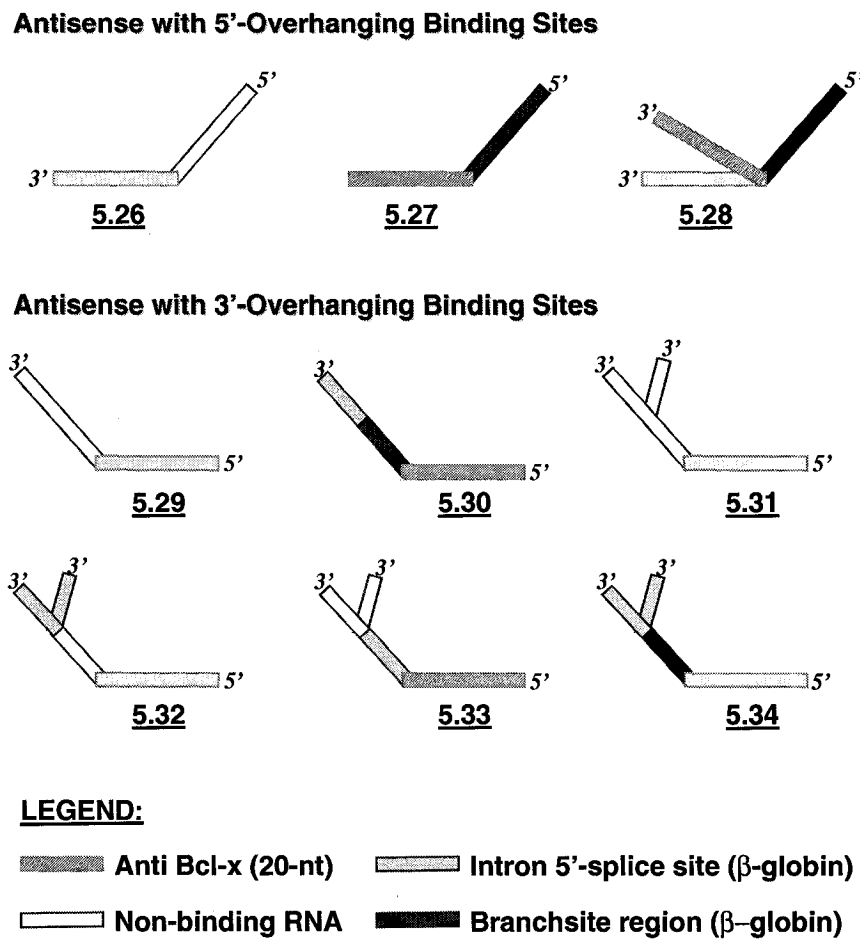


Figure 5.26: Schematic representation of the antisense, splicing factor-binding and splicing factor-non-binding regions of the oligonucleotide constructs. The 5'-splice site corresponds to a 9-nt stretch of sequences in the β-globin intron. The branchsite region represents either the full or the shortened branchsite sequences upstream in the β-globin intron.

5.4.3. *In Vitro* Modulation of Alternative Bcl-x Splicing Using Antisense Oligonucleotides Containing 5'-Overhanging Factor-Binding Sequences

The experiments associated with the modulation of Bcl-x alternative splicing using factor-binding antisense constructs were conducted in conjunction with Dr. Benoit Chabot and Jonathan Villemaire of the Université de Sherbrooke (Département de Microbiologie et d'Infectologie). As the full length Bcl-x pre-mRNA is rather long for

utilization during *in vitro* splicing studies (*i.e.* 926-nt)[136], a Bcl-x "mini-gene" (*i.e.* Bcl-x s*i*) was employed instead. This construct consists of the identical key splicing regions inherent to the full length Bcl-x. Specifically, it comprises of: (1) the 3'-end of exon 2 (371-nt) bearing two alternative splice sites for Bcl-x_L and Bcl-x_S, (2) the 5'-end of exon 3 (138-nt) containing the conserved 3'-splice site, and (3) and intronic portion (between exons 2 and 3) consisting of the first 232-nt of the 5'-region and the last 182-nt of the 3'-region in the full-length intron. The Bcl-x s*i* mini-gene was transcribed *in vitro* to its corresponding pre-mRNA by run-off transcription with T3 RNA polymerase (section 7.12.3.1). Internal radiolabels were introduced with α-[^{32}P]-UTP as a nucleoside triphosphate source. Splicing reactions were accomplished in HeLa nuclear extract as described previously.[417,450] Individual splicing reactions were supplemented with cold, exogenous antisense oligonucleotides (**Table 5.4**) at concentrations ranging from 6.6-66 nM. The product RNAs (*i.e.* alternatively spliced RNAs) were amplified by reverse transcriptase PCR (RT-PCR) to complementary DNAs (cDNAs), analyzed by non-denaturing PAGE (4%,) and the product DNAs visualized and quantitated using an InstantImager (Canberra-Packard). The amplified products Bcl-x_L (508 bp) and Bcl-x_S (319 bp) corresponded to the expected size as evidenced by their migratory behavior alongside appropriate oligonucleotide markers.

As seen in **Figures 5.27** and **Figure 5.28**, the effects of splicing modulation with both linear (**5.27**) and branched (**5.28**) oligonucleotides bearing a 5'-overhanging protein binding site were quite remarkable. In both cases, the 5'-non-complementary extension consisted of the branchsite and upstream region of the human β-globin intron 1. These non-complementary sequences would expectedly reside and obstruct the sequences downstream of the Bcl-x_L 5'-splice site. In the case of the linear construct **5.27**, the shift to predominantly Bcl-x_S formation (<20% Bcl-x_L) was apparent at 6.6 nM of antisense molecule (**Figure 5.27**; lanes 6-8 and **Figure 5.28**). At higher concentrations (*i.e.* 66 nM) negligible Bcl-x_L was observed. Similarly, the branched oligonucleotide **5.28** was also incredibly potent at redirecting splicing to the Bcl-x_S 5'-splice site (**Figure 5.27**; lanes 9-11 and **Figure 5.28**). In fact, exclusive Bcl-x_S formation was observed with this compound at the lowest concentration studied (*i.e.* 6.6 nM). This highly potent effect

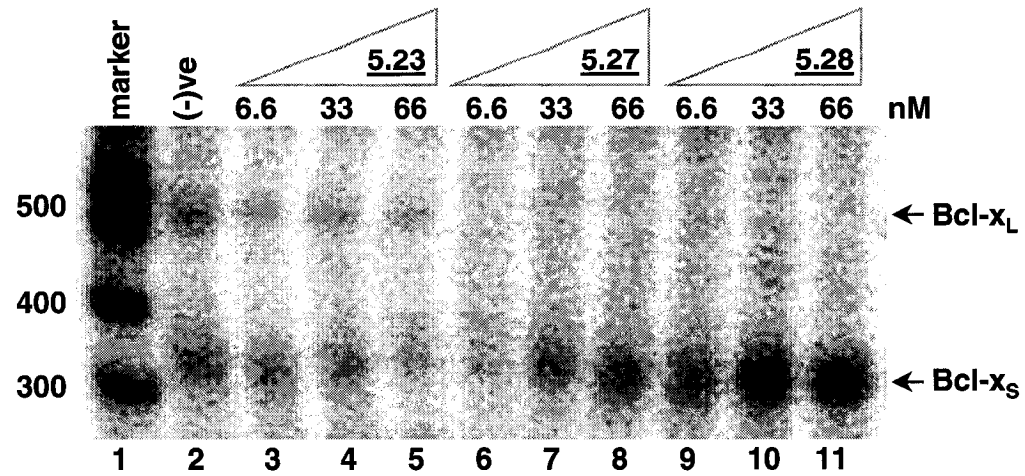


Figure 5.27: RT-PCR analysis demonstrating the modulation of Bcl-x splicing with antisense oligonucleotides containing a 5'-linear (**5.27**) or 5'-branched (**5.28**) binding region. Splicing reactions were supplemented with 6.6, 33 and 66 nM of cold antisense oligonucleotide (**5.23**, **5.27**-**5.28**). The PCR products were resolved on a non-denaturing gel (4%) and visualized by phosphorimaging. The (-)ve control is the alternatively spliced Bcl-x alone (no antisense).

may have two rationales. Firstly, **5.28** contains two complementary Bcl-x binding sequences indicating the possibility that this compound more effectively occupies the 5'-splice site of Bcl-x_L since is can recruit two pre-mRNAs for every one antisense molecule. Furthermore, elevated concentrations of antisense **5.28** increase the likeliness that the target site on any Bcl-x pre-mRNA will be occupied. Another possibility is that the branch extension is highly effective at interfering with splicing elements, thereby promoting a switch in 5'-splice site selection to that of Bcl-x_S. In contrast, antisense RNA **5.23**, which was merely complementary to the region upstream of the Bcl-x 5'-splice site, was only modestly effective at reprogramming splice site selection with a concomitant production of *ca.* 60% Bcl-x_S at all the concentrations studied (**Figure 5.27**; lanes 3-5 and **Figure 5.28**). This increase in Bcl-x_S production is very modest given the fact that the basal levels of spliced Bcl-x_L and Bcl-x_S in a negative control containing no antisense were approximately 50% each. The potency observed for **5.23** was similar to the antisense constructs previously studied by Taylor and Kole.[140,141] Additionally, the decreased activity observed with this compound may also be due to the fact that it lacked a stabilizing 3'-L-deoxynucleotide insert, and as such, was more susceptible to

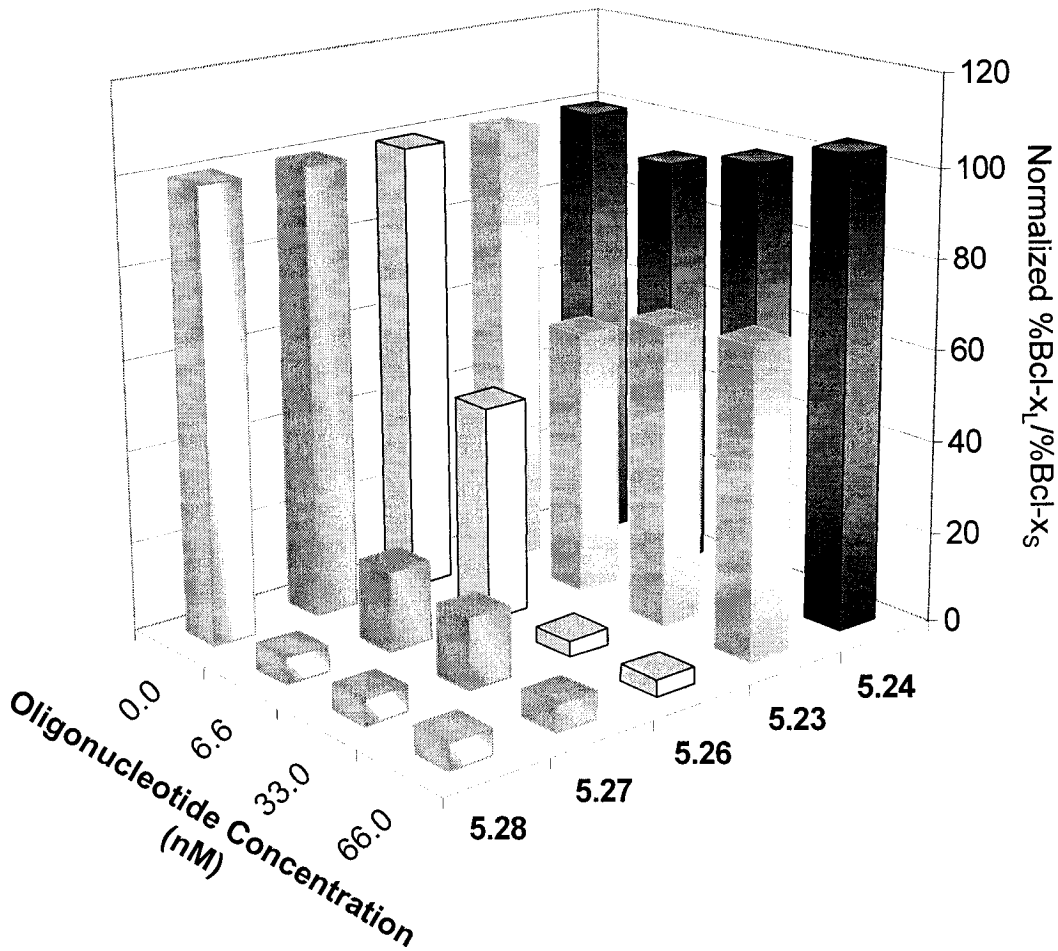


Figure 5.28: Bar graph demonstrating the efficiency of Bcl-x splicing modulation by various antisense oligonucleotides containing a 5'-overhanging binding site. The % of Bcl-x_S splicing modulation was determined by quantitation of the Bcl-x_L and Bcl-x_S spliced transcripts (see **Figure 5.27** gel), dividing the amount of Bcl-x_L by Bcl-x_S and normalizing with respect to the reaction lacking any antisense compound.

degradation by omnipresent nucleases in the HeLa extract. It is worthwhile mentioning that a construct containing the identical sequence composition to **5.23**, however incorporating nuclease-resistant 2'-O-methyl constituents, demonstrated a similar potency to **5.23** under identical splicing conditions (data not shown). In order to demonstrate that these significant observations were dependent on the dual antisense and extraneous protein-binding functionalities of the oligonucleotide constructs, a splicing modulation assay was conducted using an exogenous linear RNA (**5.24**) which merely contained the branchsite region and 5'-splice site of the β-globin intron (*i.e.* no Bcl-x complementary

region). Indeed, when a complementary antisense portion was lacking in the molecule, the amounts of Bcl-x_L and Bcl-x_S were nearly identical to the reaction lacking any exogenous oligonucleotide (**Figure 5.28**). The results suggest that the antisense constructs regulate the alternative splicing *via* a dual mode of action. Explicitly, hybridization of the antisense segment of the molecules adjacent to the 5'-splice site of Bcl-x_L impedes spliceosomal assembly at that site. This switch toward Bcl-x_S is markedly enhanced in the presence of a protein-binding oligonucleotide sequence, suggesting that such constructs interfere with splicing possibly by interfering with the initial U1 snRNP binding. Since commitment complex formation cannot occur, and hence spliceosome assembly at the Bcl-x_L 5'-splice site, this increases the frequency with which the Bcl-x_S 5' splice site is selected. To determine the sequence specificity of binding to the 5'-overhanging extension, compound **5.26** was devised. This oligonucleotide consisted of the complementary Bcl-x region as well as a control overhanging RNA tail in which the first 18-nucleotides have been shown not to bind splicing factors in HeLa nuclear extract under similar splicing conditions (Dr. B. Chabot; personal communication). At low concentration of antisense (*i.e.* 6.6 nM), the amount of Bcl-x_S splice site selection was similar to that of the antisense construct (**5.23**) lacking the interfering tail (**Figure 5.28**). Alternatively, at higher concentrations (*i.e.* 33 and 66 nM) splicing was redirected almost exclusively to the Bcl-x_S splice site. These combined results signify that the 5'-overhanging segments potentially serve a dual purpose. At lower concentrations, binding to specific sequences that are recognizable to spliceosomal factors (*i.e.* branchsites and splice sites) likely prevails, whereas at higher concentrations a steric blocking effect is the likely culprit behind the enhanced Bcl-x_S splice site selection. Alternatively, the control tail may be bound by a non-splicing factor that will interfere with U1 snRNP binding. The most potent constructs (**5.27** and **5.28**) containing the full branchsite and upstream region of the β-globin intron may potentially be bound by a branchsite recognition element such as the U2-snRNP or mBBP as described in the preceding sections. At the present time it is uncertain which and if any of these early recognition elements are recognizing the overhanging tail, and which step in the splicing reaction they may be interfering with.

5.4.4. *In Vitro* Modulation of Alternative Bcl-x Splicing Using Antisense Oligonucleotides Containing 3'-Overhanging Protein Binding Sequences

A variety of linear and branched oligonucleotide sequences containing the anti-Bcl-x segment and 3'-overhanging protein binding and non-binding sites were also developed in order to determine their effect on splicing modulation of the Bcl-x *si* pre-mRNA (**Table 5.4**). Specifically, the constructs consisted of 3'-protruding sequences that would likely reside and obstruct positions further upstream of the Bcl-x_L 5'-splice site (upstream of –24; **Figure 5.26**). Once again, the extent of 5'-splice site reprogramming was determined by quantifying the amount of Bcl-x_L and Bcl-x_S produced in the presence of cold, exogenous antisense oligonucleotides (**Figure 5.29**). As demonstrated for compound **5.30**, when the 3'-overhang consisted of a linear 21-nucleotide stretch, reminiscent of the upstream branchsite region and intronic portion of the β-globin 5'-splice site, modest modulation was observed at lower concentrations (*i.e.* 6.6 nM), however at higher concentrations (*i.e.* 66 nM), splicing resulted in exclusive Bcl-x_S formation (**Figure 5.29**; lanes 13-15 and **Figure 5.30**). Contrarily, a similar branched oligonucleotide construct, **5.34**, comprising the identical branchsite region to **5.29**, however incorporating two 5'-splice site segments was much more potent at lower

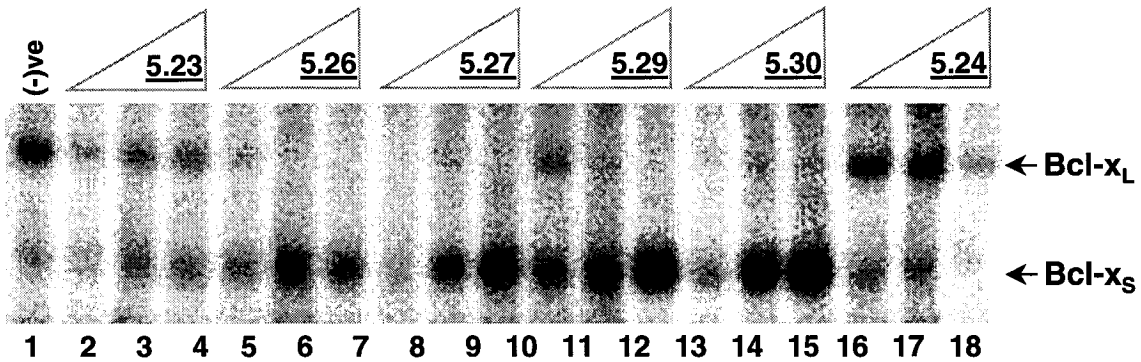


Figure 5.29: RT-PCR analysis demonstrating the modulation of Bcl-x splicing with antisense oligonucleotides containing a 5'-binding sites (**5.26** & **5.27**) or 3'-binding sites (**5.29** & **5.30**). Splicing reactions were supplemented with 6.6, 33 and 66 nM of cold antisense oligonucleotide. The PCR products were resolved on a non-denaturing gel (4%) and visualized by phosphorimaging. The (-)ve control is the alternatively spliced Bcl-x alone (no antisense).

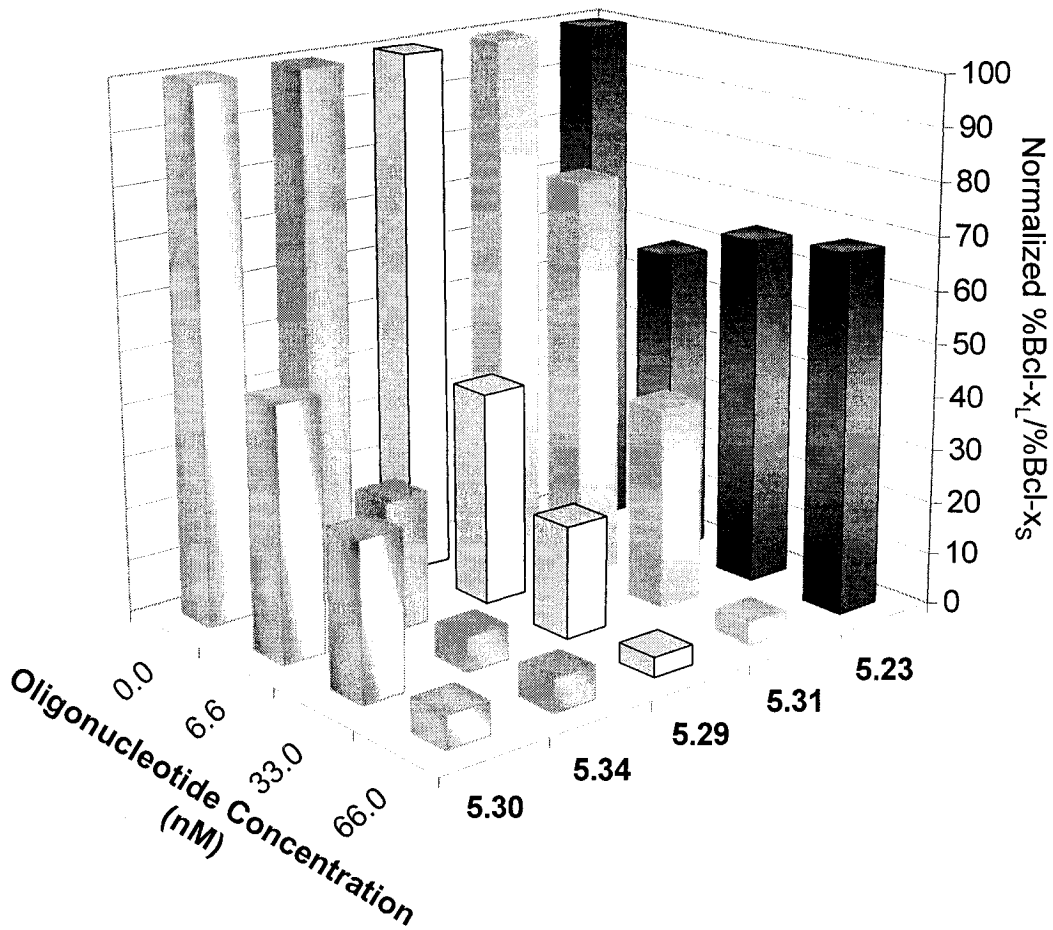


Figure 5.30: Bar graph comparing the efficiency of Bcl-x splicing modulation by various antisense oligonucleotides containing a 3'-overhanging linear and branched binding sites. The % of Bcl-x_S splicing modulation was determined by quantitation of the Bcl-x_L and Bcl-x_S spliced transcripts, dividing the amount of Bcl-x_L by Bcl-x_S and normalizing with respect to the reaction lacking any antisense compound.

concentrations (*i.e.* 83% Bcl-xS formation) and resulted in exclusive Bcl-x_S splice site selection at 33 nM (**Figure 5.30**). So as to determine the importance of the individual β-globin 5'-splice site and branchsite regions respectively on splicing modulation, branched antisense constructs **5.32** and **5.33** were created (**Table 5.4** and **Figure 5.26**). In the case of **5.32**, an 11-nucleotide stretch preceded the branchpoint and this sequence has been shown to be ineffective at binding HeLa splicing factors. Additionally, it contained a branched region in the 3'-overhang consisting of two β-globin 5'-splice sites. Compound **5.33** on the other hand, contained the splice site preceding the branchpoint and the control, non-binding RNA sequences off the 2' and 3'-positions of the branchpoint in the

overhang. As seen in **Figure 5.31**, both the branched constructs **5.32** and **5.33** were nearly equally effective at transmitting 5'-splice site selection towards the Bcl-x_S 5'-splice site at low concentration of oligonucleotide (*i.e.* 60-70% Bcl-x_S formation), but demonstrated a slightly less potent effect than **5.34**. Nonetheless, at higher concentrations, Bcl-x_L production was completely eliminated by both oligomers. Comparatively, supplementation of the splicing reactions with exogenous **5.31**, which did not contain any protein binding sites revealed a drastic decrease in Bcl-x_S splicing levels at both 6.6 and 33 nM, but exclusive Bcl-x_S formation at 66 nM. In contrast, linear antisense **5.29**, which contained the control RNA tail (*i.e.* non-splicing factor binding) appeared to be just as effective as its linear derivative **5.30** at all the concentrations studied (**Figure 5.30**). Moreover, when the bNA **5.25**, which lacked an anti-Bcl-x sequence, was analyzed for its potential to redirect splicing, the results undoubtedly indicated that splicing modulation strictly required the presence of a complementary 5'-splice site oligonucleotide, since this molecule was completely ineffective at altering splice site selection (**Figure 5.31**). These results clearly suggest an important role for the factor-binding sites in the branched antisense constructs for effective modulation of Bcl-x splicing. The sequences encompassing the branchpoint appear to be of greater interfering ability at lower concentrations than at higher ones, implying that branch recognition by a spliceosomal element may be more interfering (compare **5.31**-**5.34**; **Figure 5.31**). The highest potency observed with **5.34** evidently implies that both a functional 5'-splice site and branchsite region in the overhang contribute to potency. The difference in activity between **5.34** and those lacking the branchsite region in the overhang (**5.32** and **5.33**) may point towards a difference in the interference step by various spliceosomal elements. A 5'-splice site overhang may compete with initial recognition events by an early spliceosomal element such as the U1-snRNP. Contrarily, the branchsite overhang may interfere with both early and late steps of the spliceosomal assembly process through recognition by the U2-snRNP, since it is well established that this element makes direct contact with both a pre-branched and branched *(i.e.* lariat-RNA) RNA at both steps of the splicing reaction.[111] Given that the U2 and U1 snRNPs are large molecular complexes, we can rationalize how their binding in the vicinity of a 5'-splice site can still potently interfere with normal spliceosome assembly even when their binding is centered at

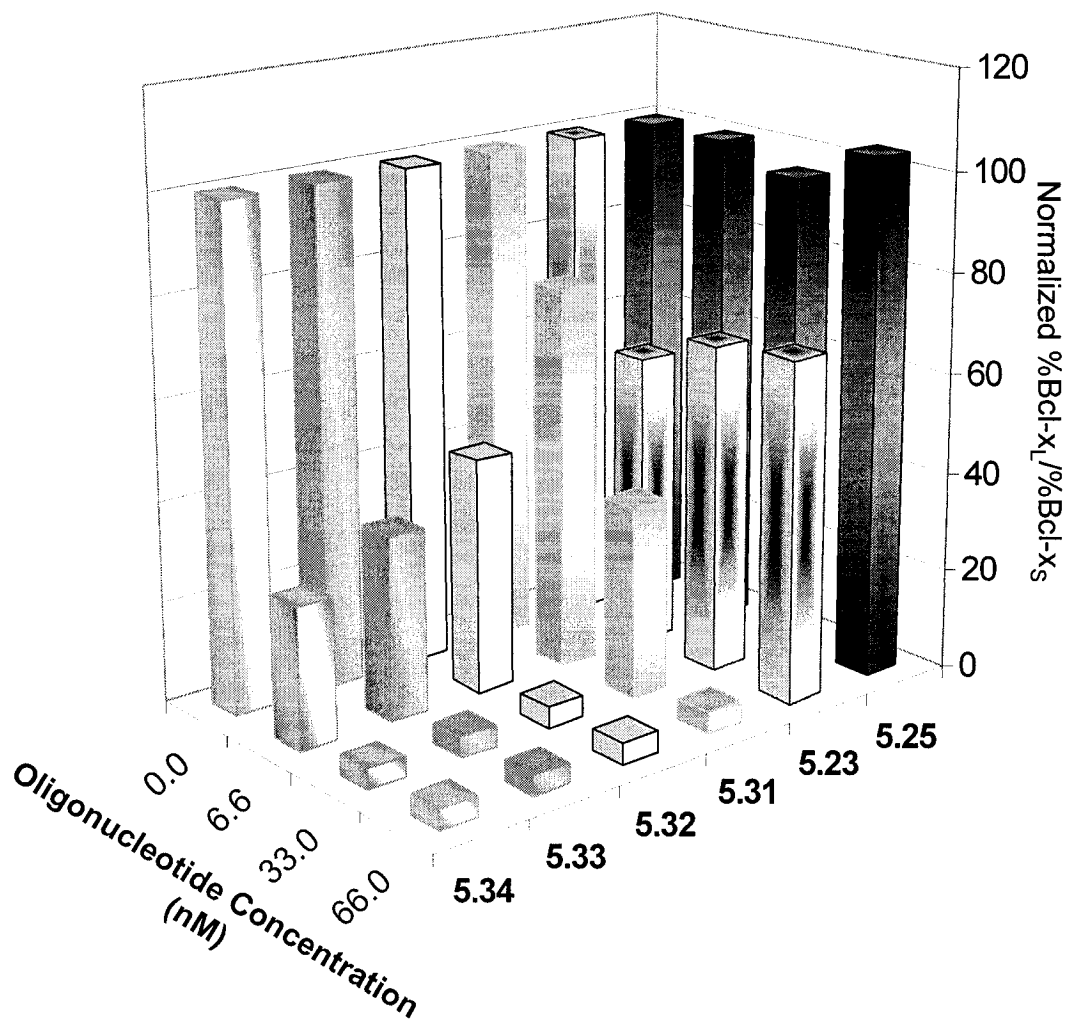


Figure 5.31: Bar graph demonstrating the efficiency of Bcl-x splicing modulation by various antisense oligonucleotides containing a 3'-overhanging branched binding sites. The % of Bcl-x_S splicing modulation was determined by quantitation of the Bcl-x_L and Bcl-x_S spliced transcripts, dividing the amount of Bcl-x_L by Bcl-x_S and normalizing with respect to the reaction lacking any antisense compound.

position –24 relative to the 5'-splice site. The high activity seen with the control sequences **5.29** and **5.31** at elevated concentrations may also point towards late recognition by an unknown factor, which may be less interfering during early spliceosome assembly or simply less abundant. To decipher which factors and at what step the interference in occurring, biochemical analysis of the elements making direct contact with the overhanging regions in the antisense constructs is necessary. Given the

profusion of information we have obtained on branch recognition using synthetic branched oligonucleotide constructs, their behaviors in nuclear extracts, and the variety of methods we can use to stabilize these constructs against degradation by ubiquitous nucleases and debranching activity, we are now in a position to assess their complete biological function *in vivo*.

5.5. CONCLUSIONS

The preliminary results obtained in the preceding section clearly point toward the eminent utility of antisense oligonucleotides containing factor-binding and/or steric interfering overhangs for the effective modulation of Bcl-x splice site selection. Specifically, the effects of Bcl-x_S splice site selection appear to be superior with oligonucleotides containing a branched core linked by both 2',5' and 3',5-phosphodiester bonds to a specific protein or spliceosomal factor recognition element such as an intact branchsite or 5'-splice site. The most potent modulator comprising a 5'-overhang was compound **5.28**; a branched RNA bearing two anti-Bcl-x binding regions and a β-globin upstream branchsite. Whether this construct exhibits its high potency due to the fact that it occupies more Bcl-x pre-mRNAs or more effectively sterically hinders splicing remains unknown at this time. Additionally, the most potent molecule containing a 3'-overhang was the branched **5.34**, suggesting that an intact branchsite and β-globin 5'-splice site may be key recognition motifs possibly contributing to both early and late stage interference during the splicing reaction. In order to ascertain which elements may be contributing to exogenous RNA recognition during alternative splicing of the Bcl-x pre-mRNA, we are currently pursuing a biochemical investigation of the factors that conceivably make direct contact with the 5'- or 3'-overhangs present in the antisense agents. Furthermore, we are presently conducting additional *in vitro* assays at lower concentrations of antisense construct in order to achieve a reasonable limit of confidence with respect to differences in splicing modulation effects.

Overall, we are highly encouraged by the exceptionally promising results obtained. By taking advantage of the fact that the alternative splicing of Bcl-x results in the conception of two protein isoforms (Bcl-x_L and Bcl-x_S) with antagonistic functions in terms of oncogenic cell survival, we have shown that our strategy may potentially be useful as a therapeutic agent for the control of cancerous cell proliferation. Similar constructs, however lacking the protein-binding extensions, have already proven successful for the amplification of the pro-apoptotic effects of Bcl-x_S in lung[141], breast and prostate[140] cancer cell lines, however the expression levels of the Bcl-x_S isoform vary among the cell lines tested and in the majority of cases required adjuvant treatment with a cytotoxic agent. As such, the markedly increased shift in Bcl-x_S spliced variant obtained with our constructs may result in amplified expression of this isoform across cell lines, thereby leading to potential full exploitation of the pro-apoptotic potential of this protein. There is also increasing and convincing evidence that such antisense oligonucleotides, targeted toward specific RNA sequences, exert their biological effects predominantly, if not exclusively, in the cell nucleus.[451] Knowing this, the antisense-mediated modulation of Bcl-x alternative splicing should prove highly alluring given that pre-mRNA splicing is localized in the nuclear compartments of a cell.

CHAPTER 6: CONTRIBUTIONS TO KNOWLEDGE

6.1. CONCLUSIONS AND FUTURE WORK

A novel synthetic methodology towards the creation of lariat-DNA structures on solid-support by exploiting the differential rates of cleavage of two CPG-tethers, namely the hydroquinone-O,O'-diacetate (*Q-linker*) and succinate (*S-linker*) moieties has been demonstrated. This synthesis takes advantage of our well-established convergent strategy for the assembly of branched DNA molecules on solid-support using the bifunctional bis-adenosine phosphoramidite synthon, which effectively reacts with two adjacent support-bound linear oligonucleotide strands, thus forming a branch juncture. Additionally, a new phosphodiester bond can be successfully introduced by first phosphitylating the 5'-hydroxyl of a CPG-linked nucleoside or oligonucleotide followed by activation of the phosphoramidite moiety and coupling to the 3'-hydroxyl of an exogenous nucleoside or intramolecular DNA sequence. An absolute correlation between CPG-loading and high branching efficiency has also been established, which consequently reveals itself to be a major deterrent for effective cyclization of support-bound branched DNA molecules, thereby leading to predominant synthesis of multi-branched dendrimer species. A small amount of cyclized lariat product can nonetheless be isolated from the complex reaction mixture and its topology secured by a variety of analytical and enzymatic techniques such as denaturing PAGE, HPLC and hydrolysis with the 2'-scissle activity found in HeLa nuclear extracts. Regioselectivity in the cyclization reaction cannot however be afforded using this method as both the 2' and 3'-tethered oligonucleotide strands are concomitantly released upon mild basic treatment of the *Q-linker*. Furthermore, in order to reduce the entropic penalty of cyclizing two freely rotating oligonucleotide strands, it may be worthwhile to construct shorter oligonucleotide species such that intramolecular cyclization can be more favorably driven.

In an effort to devise of a more facile, versatile and regioselective approach to lariat olignucleotide synthesis, specifically for the creation of the more biologically relevant RNA lariat, the template-mediated construction of branched, cyclic oligonucleotides was

investigated. The results demonstrate the first example of a simplistic approach to lariat-oligonucleotide synthesis, and exploit the unique topology of synthetic branched nucleic acids. This effective strategy necessitated the use of only commercially available phosphoramidite building blocks except for the adenosine-bis-phosphoramidite, which can be readily prepared in-house by previously published and accepted methods. Once again, cyclization of the lariat species relied on the use of convergently and divergently synthesized branched DNA or RNA molecules as precursors, which are simple, efficient and inexpensive to prepare by any laboratory self-sufficient in automated oligonucleotide synthesis. Both intramolecular templates (*i.e.* dumbbell structures), and intermolecular splints (*i.e.* bimolecular association) are equally effective at producing high yields of cyclized lariat structures upon chemical condensation of a phosphate/hydroxyl impingement with an inexpensive reagent such as cyanogen bromide. Specifically, the yields of medium-sized DNA and RNA lariats that were 46-57 nucleotides long, were in the 40-60% range starting from a branched nucleic acid precursor. Regioselectivity in cyclization is clearly afforded when the precursors consist of diverse 2'- and 3'-extensions off the branch, as in the case of the DNA lariats. Once an effective strategy for the regiospecific synthesis of branched RNA molecules is accomplished in the solid-phase, these precursors will be readily amenable to the identical regioselective cyclization reactions described for DNA, thereby affording RNA lariats that are truly reminiscent of the native biomolecules. Additionally, it was clearly shown that lariat molecules can be characterized by a variety of methods including denaturing PAGE, CD and T_m analysis, as well as enzyme-based systems such as the 5'-exonuclease, BSPDE.

Given the remarkable potencies obtained with RNA hairpin molecules for the specific inhibition of the RNase H activity of HIV-1 reverse transcriptase in our laboratory, it was proposed that RNA dumbbell molecules, containing a double-hairpin like structure, might yield the same or even more powerful effects. Specifically, such dumbbell structures were particularly interesting given the fact that they display remarkable nuclease resistance, owing to the fact that they bear no free 3'-ends. High yielding synthesis of the RNA dumbbells is very easily effected by chemical condensation of a self-complementary RNA oligonucleotide with CNBr. The RNA dumbbells demonstrate a

remarkably selective effect towards the RNase H activity of HIV-1 RT without any consequence on its polymerase functionality. This was most notably seen by a UV-crosslinking assay with the RNase H domain of the enzyme. Importantly, non-specific inhibition of either the RNase H activities from bacterial (*i.e.* E. coli) or mammalian (*i.e.* human) was not observed, indicating an extraordinary specificity of such structures towards the retroviral enzyme only. The most potent construct, and RNA dumbbell containing a ten base-pair stem and two flanking UUCG loop motifs, displayed an IC_{50} in the 3 μM range. The results further support the fact that the UUCG loop as well as the length of the stem in both hairpin and dumbbell structures are important factors for achieving high RNase H inhibitory potency. The stability, delivery and activity of this novel class of oligonucleotide-based inhibitors in HIV-infected cells will ultimately determine if such constructs can indeed be developed into effective anti-retroviral therapeutics.

Branched RNA oligonucleotides have demonstrated great utility as agents for the study of branchpoint recognition during nuclear pre-mRNA splicing in both yeast and mammalian systems. Stabilization of such constructs against pervasive endonuclease and exonuclease activity is requisite for achieving notable splicing inhibition. The incorporation of one unnatural L-deoxynucleotide at the 3'-terminus of an otherwise unmodified oligonucleotide was shown to demonstrate adequate resistance toward exonuclease hydrolysis. Furthermore, replacement of the natural ribopyrimidine residues by 2'-OMe ribopyrimidines produced structures that were even more resistant towards nucleases, presumably of the endonuclease type. Additionally, more enhanced inhibitory potential is seen with this modification and suggests that the 2'-OMe inserts promote tighter binding to the spliceosomal element(s) which specifically recognize the branched architecture. Inhibitory activity is also maintained in a bNA containing an arabinoadenosine branchpoint, which was also shown to be completely resistant to the 2'-phosphodiesterase activity (*i.e.* debranching) present in the biological milieu. Overall, a comprehensive evaluation of the requisite features for stabilizing branched oligonucleotides against enzymatic activities present in extracts has been established. The global results put us in a very promising position to conduct a full biochemical analysis of the factor(s) involved

in branch recognition during pre-mRNA splicing, since this phenomenon is not well understood at the present time. By introducing photoreactive moieties into appropriately stabilized branched RNA molecules during solid-phase oligonucleotide assembly, crosslinking to the putative branch recognition is now possible. In conjunction, affinity selection of the spliceosomal or non-spliceosomal element(s) can also be attempted, thereby yielding more conclusive evidence as to the nature of this recognition event during splicing.

Our studies on the modulation of Bcl-x alternative splicing used linear and branched antisense oligonucleotides saw the syntesis of the longest bRNA structures ever realized by any laboratory. Most importantly, we have demonstrated how antisense oligonucleotides containing protein-binding overhangs can effectively alter splicing of Bcl-x from the anti-apoptotic Bcl-x_L to the pro-apoptotic Bcl-x_S isoform. A dual mode of activity is presumed in these compounds whereby the antisense part of the molecule redirects spliceosomal assembly from the 5'-splice site of Bcl-x_L towards that of Bcl-x_S, and the protein-binding overhangs act as steric interfering agents by blocking spliceosomal assembly events or sequestering necessary spliceosomal elements. The levels of Bcl-x_S obtained were the highest ever reported by any group, and may potentially lead to higher expression levels of Bcl-x_S in cell-based systems. This is most important in cell lines that appear to express lower levels of Bcl-x_S such as lung and breast cancers.

Overall, the utility of linear, branched and circular oligonucleotides for construction of higher-ordered oligonucleotide species (*i.e* lariats), as well as agents for the control of splicing, retroviral activity and gene expression has been suitably demonstrated. The future looks even more promising.

6.2. CONTRIBUTIONS TO KNOWLEDGE

As a direct result of the studies described herein, the following publications have recently emerged, have been submitted for publication or are currently in preparation:

(1) Jonathan Villemaire, Sandra Carriero, Benoit Chabot and Masad J. Damha. "*In vitro* Splicing Modulation of Bcl-x Alternative Splicing by Branched Oligonucleotides" (in preparation).

(2) Sandra Carriero and Masad J. Damha. "Inhibition of pre-mRNA Splicing by Branched Oligonucleotides", *Nucleic Acids Research* **2003,** *31*, 21, 6157-6167.

(3) Rami N. Hannoush, Sandra Carriero, Kyung-Lyum Min and Masad J. Damha. "Selective Inhibiton of HIV-1 Reverse Transcriptase (HIV-1 RT)RNase H by Small RNA Hairpins and Dumbbells", *ChemBioChem* **2004** (in press).

(4) Sandra Carriero and Masad J. Damha. "Template-Mediated Synthesis of Lariat-RNA and DNA", *Journal of Organic Chemistry* **2003,** *68*, 22, 8328-8338.

(5) Sandra Carriero and Masad J. Damha. "Synthesis of Lariat-DNA *via* the Chemical Ligation of a Dumbbell Complex", *Organic Letters* **2003,** *5*, 273-276.

(6) Sandra Carriero and Masad J. Damha. "Solid-Phase Synthesis of Branched Oligonucleotides". In *Current Protocols in Nucleic Acids Chemistry*; Beaucage, S.L. Ed, John Wiley & Sons: New York, Unit 4.14, pp. 1-32.

(7) Sandra Carriero, Ravi S. Braich, Robert H.E. Hudson, Dorian Anglin, James D. Friesen, and Masad J. Damha. "Inhibition of *in vitro* pre-mRNA Splicing in *S. cerevisiae* by Branched Oligonucleotides", *Nucleosides, Nucleotides and Nucleic Acids* **2001**, 4-7, 873-877.

Contributions to multi-authored work (numbers correspond to publication list above):

(1) "*In vitro* Splicing Modulation of Bcl-x Alternative Splicing by Branched Oligonucleotides"

Carriero synthesized, purified and analyzed all the linear and branched oligonucleotide constructs reported in this study. Bcl-x alternative splicing modulation assays were conducted by Jonathan Villemaire in the laboratory of Dr. Benoit Chabot at the Université de Sherbrooke (Département de Microbiologie et Infectologie).

(3) "Selective and Potent Inhibiton of HIV-1 Reverse Transcriptase RNase H by Small RNA Hairpins and Dumbbells"

Carriero contributed to all work concerning the chemical synthesis, purification, analysis of the RNA dumbbells described.

(6) "Inhibition of *in vitro* pre-mRNA Splicing in *S. cerevisiae* by Branched Oligonucleotides"

Carriero synthesized, purified and characterized all the 3'-L-deoxynucleotide stabilized linear, V-shaped and Y-shaped oligonucleotides. Furthermore, Carriero contributed to the preparation of yeast nuclear extract and the actin pre-mRNA transcript during a visit to the laboratory of Dr. James D. Friesen at the University of Toronto (Banting and Best Department of Medical Research) as well as conducting preliminary *in vitro* pre-mRNA splicing optimization assays.

On account of the results disclosed in Chapter 4, the following provisional patent is currently being compiled:

Masad J. Damha and Sandra Carriero (inventors). "Selective and Potent Inhibition of HIV-1 Reverse Transcriptase RNase H by an RNA Dumbbell", Provisional U.S. Patent, in preparation.

Publications resulting from studies not reported in this thesis:

(1) Sandra Carriero, Maria Mangos, Kazim Agha, Anne Noronha and Masad J. Damha. "Branchpoint sugar stereochemistry determines the hydrolytic susceptibility of branched RNA fragments by yDBR", *Nucleosides, Nucleotides and Nucleic Acids* **2003** *22*, 5-8, 1599-1602.

Aspects of the studies described in this thesis have been presented at the following:

Oral Presentations (underlined name represents presenting author):

(1) Sandra Carriero and Masad J. Damha; Inhibition of Mammalian pre-mRNA Splicing Using Synthetic Branched Nucleic Acids. **Award Lecture for Best Student Presentation in 2002-2003**; **Eastern Ribo-Club Meeting**, Mont-Orford, Quebec, September 2002.

(2) Sandra Carriero and Masad J. Damha; Template-Mediated Chemical Synthesis of Nucleic Acid Lariats. **Eastern Ribo-Club Meeting**, University of Sherbrooke, Sherbooke, Quebec, January 2003

(3) Sandra Carriero and Masad J. Damha; Chemical Synthesis of Oligonucleotide Lariats, **85th Annual CSC Conference and Exhibition**, Vancouver, BC, June 2002.

(4) Maria Mangos, Sandra Carriero, Kazim Agha, Anne Noronha and Masad J. Damha; Substrate Specificity of the Yeast Lariat Debranching Enzyme. **85th Annual CSC Conference and Exhibition**, Vancouver, BC, June 2002.

(5) Sandra Carriero and Masad J. Damha; Forks, Dumbbells and Lassos: Exploiting the Topology of Branched DNA for the Synthesis of Lariat Oligonucleotides. **McGill University Organic Seminar Series**, Montreal, QC, April 2002.

(6) Sandra Carriero, Kanjana Ganeshan, Ravindjervit S. Braich and Masad J. Damha; Attempts Towards the Solid-Phase Synthesis of Lariat Oligonulceotides. **84th Annual CSC Conference and Exhibition**, Montreal, Quebec, May 2001.

(7) Sandra Carriero, Dorian Anglin, Ravindjervit S. Braich, James D. Friesen and Masad J. Damha; Synthetic Branched Oligonucleotides Inhibit *in vitro* pre-mRNA Splicing in *S. Cerevisiae*, **Eastern Ribo-Club Meeting**, University of Sherbrooke, Sherbrooke, Quebec, February 2001.

Poster Presentations (underlined name represents presenting author):

(1) Sandra Carriero and Masad J. Damha; An Intramolecular Template Approach to Lariat Oligonucleotide Synthesis, **Eastern Ribo-Club Meeting**, Mont-Orford, Quebec, September 2002.

(2) Sandra Carriero and Masad J. Damha; Synthesis of Small "Lariat" Oligonucleotides: A Novel Solid Phase Approach, **Eastern Ribo-Club Meeting**, Mont-Orford, Quebec, September 2001.

(3) Sandra Carriero and Masad J. Damha; Novel Synthesis of a DNA Lariat on Solid Support, **Concordia University 4th Annual Chemistry and Biochemistry Graduate Research Conference**, Montreal, Quebec, September 2001.

(4) Sandra Carriero and Masad J. Damha; Probing Branch Point Recognition During pre-mRNA Splicing Using Synthetic Branched Oligonucleotides, **84th Annual CSC Conference and Exhibition**, Montreal, Quebec, May 2001.

(5) Sandra Carriero, Dorian Anglin, Ravindjervit S. Braich, James D. Friesen and Masad J. Damha; Inhibition of *in vitro* pre-mRNA Splicing in *S. Cerevisiae* by Branched Oligonucleotides, **XIV International Roundtable: Nucleosides, Nucleotides and Their Biological Applications**, San Francisco, California, U.S.A., September 2000.

(6) Sandra Carriero, Dorian Anglin, Ravindjervit S. Braich, Masad J. Damha, and James D. Friesen; Branched RNA Recognition by a pre-mRNA Splicing Factor, **Concordia University 3rd Annual Chemistry and Biochemistry Graduate Research Conference**, Montreal, Quebec, March 2000.

(7) Sandra Carriero, Dorian Anglin, Masad J. Damha and James D. Friesen; Inhibition of yeast pre-mRNA splicing by synthetic branched oligonucleotides, **10th Annual Quebec-Ontario Minisymposium in Synthetic and Bio-Organic Chemistry**, Saint-Sauveur des Monts, Quebec, November 1999.

CHAPTER 7: EXPERIMENTAL MATERIALS AND METHODS

7.1. GENERAL REAGENTS

Anhydrous reagents for reactions and oligonucleotide synthesis were either purchased from commercial sources or prepared in-house according to known methods. Acetonitrile (CH_3CN) was pre-dried by refluxing over phosphorus pentoxide (P_2O_5; Fisher) followed by reflux and distillation over calcium hydride (CaH_2; Aldrich) under inert atmosphere. The dry acetonitrile was distilled just prior to use and withdrawn from the septum port of a collection bulb *via* dry syringe. Tetrahydrofuran (THF) was dried by continuous reflux over sodium metal and benzophenone under an inert atmosphere until a purple color persisted, distilled into a collection bulb and withdrawn from the septum port *via* a dry syringe. Pyridine (Fisher), collidine (2,4,6-trimethylpyridine; Aldrich), triethylamine (Fisher) and N-methylimidazole (NMI; Aldrich) were dried by refluxing and distillation over CaH_2 and stored under inert atmosphere over activated molecular sieves. N,N-diisopropylethylamine (DIPEA; Aldrich) was dried by mild heating and stirring over CaH_2 (50-60°C, 16h) followed by distillation under vacuum and stored over activated molecular sieves in septum-sealed bottles. Acetic anhydride and reagent grade 1,2-dichloroethane (DCE) were obtained from Fisher and used as received.

1-H-tetrazole, O-(1H-Benzotriazol-1-yl)-1,1,3,3-tetramethyl-uronium hexafluorophosphate (HBTU) and O-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyl-uronium hexafluorophosphate (HATU) were obtained from Applied Biosystems Inc. (Foster City, CA) and stored in a dessicated environment at room temperature. 4,5-dicyanoimidazole (DCI, 99%; GLS Synthesis Inc., Worcester, MA) and 1-H-tetrazole (sublimed, 99%) were dried under vacuum over P_2O_5 for 24 h preceding use. N-cyanoimidazole (ImCN) was purchased from Toronto Research Chemicals and stored at –20°C under argon. β-Cyanoethyl-(N,N-diisopropylamino) phosphochloridite was purchased form ChemGenes Corp. (Ashland, MA) and stored at –20°C. It was warmed to room temperature in a dessicated environment before use.

4-(Dimethylamino)-pyridine (4-DMAP), 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide (DEC), 4,4'-dimethoxytriphenylmethyl chloride (dimethoxytrityl chloride, DMT-Cl), benzoyl chloride, 1M tetra-*n*-butylammonium fluoride (TBAF) in THF, trichloroacetic acid (TCA), 2-(N-morpholino)ethanesulfonic acid (MES) and 5M cyanogen bromide (CNBr) in acetonitrile were all purchased from Sigma-Aldrich Canada.

7.2. General Methods

7.2.1. Chromatography

Thin layer chromatography (TLC) was performed on Merck Kieselgel 60 F_{254} silica gel-coated aluminum sheets with fluorescent indicator (0.2 mm × 20 cm × 20 cm). Visualization of compounds was conducted by either illumination with a handheld UV lamp at 254 nm and/or by dipping the TLC plate in a sulfuric acid-based solution in order to visualize any trityl-bearing species.

Purification of compounds by flash column chromatography was conducted on silica gel (230-400 mesh, Silicycle Inc., QC), using 20-25 g of silica gel per gram of crude material.[452]

7.2.2. Instrumentation

UV-Vis spectra. UV and visible absorbance spectra were collected on a Varian Cary 1 UV/Vis spectrophotometer (Mulgrave, Victoria, Australia) equipped with a multiple cell holder and a Peltier temperature controller. Unless otherwise noted, all spectra were obtained in water or aqueous buffers. The data obtained was analyzed using software supplied by the manufacturer (Cary Win UV, version 2.00). Investigation of oligonucleotide complexes was conducted by thermal denaturation analysis (Section 7.6.1)

Fast Atom Bombardment Mass Spectrometry (FAB-MS). Mass spectrometric analysis was conducted on a Kratos MS25RFA high-resolution mass spectrometer using a *p*-nitrobenzyl alcohol (NBA) matrix.

7.3. General Automated Solid-Phase Oligonucleotide Synthesis

7.3.1. General Reagents

Reagents for the solid-phase synthesis of oligonucleotides were prepared from the highest quality synthesis grade materials. All solvents (*e.g.* acetonitrile, THF) were dried under an inert atmosphere as described previously (*vide supra*), distilled immediately prior to use and withdrawn from a septum-sealed collection bulb with a dry glass syringe and stainless steel needle. The phosphoramidite activation reagent consisted of either 0.5 M tetrazole or 0.5 M 4,5-dicyanoimidazole in anhydrous acetonitrile. Removal of the 5'or 3'-dimethoxytrityl groups was carried out with a solution of 3% (w/v) trichloroacetic acid (TCA) in 1,2-dichloroethane (DCE). The reagents for the acetylation or "capping" of any unreacted hydroxyl groups bound to the solid-support were prepared as follows: Cap A; 10% (v/v) acetic anhydride and 10% (v/v) dry 2,4,6-collidine in anhydrous THF, Cap B; 16% (v/v) dry NMI in anhydrous THF. Intermediate phosphite triesters were oxidized to the more stable phosphotriesters using a solution of 0.1 M iodine (Aldrich) in THF/pyridine/water (25:20:2, v/v/v).

With the exception of the convergent branching phosphoramidite monomer, 5'-O-dimethoxytrityl-(N^6-benzoyl adenosine)-2', 3'-O-bis-(β-cyanoethyl) N,N-diisopropyl phosphoramidite (Compound **1.1**; Section 7.8.5), all standard 3'-DNA phosphoramidites [5'-O-dimethoxytrityl-2'-deoxy-(N-protected ribonucleoside)-3'-O-(β-cyanoethyl) N,N-diisopropyl phosphoramidites], inverted 5'-DNA phosphoramidites [3'-O-dimethoxytrityl-2'-deoxy-(N-protected ribonucleoside)-5'-O-(β-cyanoethyl) N,N-diisopropyl phosphoramidite], 3'-RNA phosphoramidites [5'-O-dimethoxytrityl-2'-O-

tert-butyldimethylsilyl-(N-protected ribonucleoside)-3'-O-(β-cyanoethyl) N,N-diisopropyl phosphoramidite] and 2'-RNA phosphoramidites [5'-O-dimethoxytrityl-3'-O-*tert*-butyldimethylsilyl-(N-protected ribonucleoside)-2'-O-(β-cyanoethyl) N,N'-diisopropyl phosphoramidite] were purchased from ChemGenes Corp (Ashland, MA), stored at –20°C and dried *in vacuo* overnight over P_2O_5 prior to use. The N^6- and N^4-amino groups of adenine (Ade) and cytosine (Cyt) were protected with benzoyl (Bz) and N^2-amino of guanine (Gua) protected with an isobutyryl (*i*-Bu) group. The 3' and 5' chemical phosphorylation reagent: 2-[2-(4,4'-dimethoxytrityloxy)ethylsulfonyl]ethyl-(β-cyanoethyl)-(N,N-diisopropyl) phosphoramidite was purchased from ChemGenes and stored at –20°C.

Long chain alkylamine controlled-pore glass (LCAA-CPG; 500 Å or 1000 Å pore size) was obtained from Dalton Chemical Laboratories (Toronto, ON) or CPG Inc. (Lincoln Park, NJ) and derivatized with the appropriate linker-arm and nucleoside as described below (section 7.3.3). Suitably protected 5'-O-dimethoxytrityl-N-protected-2'-deoxyribonucleosides or 5'-O-dimethoxytrityl-N-protected-ribonucleosides bearing free 3'-hydroxyl termini for esterification to succinylated CPG were obtained from ChemGenes Corp.

7.3.2. Solid-Phase Synthesis of Oligonucleotides

Synthesis of linear and branched oligonucleotides was conducted on an Applied Biosystems (ABI) 381A synthesizer using the standard β-cyanoethyl phosphoramidite chemistry and the 1 μmol scale delivery cycle supplied by the manufacturer with slight modifications. The standard cycle was customized to include the following: (1) *phosphoramidite coupling*: the coupling time or "wait" step was extended to 120 seconds for the 2'-deoxyribonucleoside phosphoramidites (dA, dC, dT) and 240 seconds for dG, as well as 450 seconds for the ribonucleoside phosphoramidites (rA, rC, rU) and 600 seconds for rG, (2)*capping*: acetylation of unreacted 5'-hydroxyl groups was accomplished by a 17 second delivery of Cap A and Cap B reagent followed by a 45 second "wait" step and repeated, (3)*oxidation*: oxidant solution was delivered to the

column for 20 seconds followed by a 20 second "wait" step (4)*detritylation*: solution of 3% TCA in DCE was continuously delivered for 120 seconds to the column for the removal of DMT-containing groups.

Preceding the assembly of the oligonucleotide chain, the nucleoside-derivatized solid support (1 μmol CPG) was packed into an empty synthesizer column (ABI), installed on the instrument and treated with a mixture of Cap A and Cap B reagents according to the pre-installed "capping" cycle provided by ABI. This step ensures that any undesired hydroxyl or amino groups on the CPG surface are masked by acetylation[166,178], and also eliminates any trace moisture at the beginning of the synthesis. Phosphoramidite reagents were dissolved in freshly distilled acetonitrile, which was introduced *via* dry syringe through the septum of a sealed amber glass bottle containing the appropriate monomer. The final concentrations of the conventional monomers were 0.1 M for the 2'-deoxyribonucleoside phosphoramidites and 0.15 M for the ribonucleoside phosphoramidites unless otherwise noted. Working concentrations of any atypical nucleoside and non-nucleoside phosphoramidites are defined in their appropriate sections. Monitoring of successive coupling efficiencies was conducted by measuring the absorbance (at 505 nm) of the trityl cation released during the TCA treatment step.

7.3.3. Derivatization of Controlled Pore Glass (CPG) Supports

7.3.3.1. Acid Activation[165] and Succinylation of the CPG Support [166]

LCAA-CPG (4 g, 500 Å or 1000 Å pore size) was treated with a solution of 3% TCA in DCE for 24-48 h at room temperature in order to liberate a maximal number of reactive amino sites on the support surface. The activated CPG was filtered and neutralized by washing with 9:1, triethylamine:DIPEA (50 mL), washed successively with dichloromethane and diethyl ether, and placed in a vacuum dessicator to dry (12-24 h) prior to succinylation.

The acid-activated LCAA-CPG (1 g), succinic anhydride (2 mmol, 0.20 g) and 4-DMAP (0.33 mmol, 40 mg) were placed in a septum-sealed 10 mL glass vial. Anhydrous

pyridine (6 mL) was added *via* syringe and the vial shaken gently at room temperature for 24 h. The contents were filtered and washed sequentially with pyridine, dichloromethane and diethyl ether and placed in a vacuum dessicator over phosphorus pentoxide to dry.

7.3.3.2. Nucleoside Derivatization of CPG

Method A (Average Loading: 25-50 μmol/g).

CPG loadings of all deoxyribonucleoside and ribonucleosides that are acceptable for the synthesis of linear oligonucleotides (<50 nts = 500 Å CPG, >50 nts = 1000 Å CPG) can be attained using the coupling reagents DEC and DMAP according to the method of Damha *et al.*[166] Briefly, succinylated LCAA-CPG (0.5 g), 5'-O-DMT-N-protected nucleoside (0.1 mmol), 4-DMAP (0.05 mmol, 6 mg) and DEC (1.0 mmol, 192 mg) were placed into a septum-sealed 10 mL glass vial. Anhydrous pyridine (6 mL) and triethylamine (40 μL) were added and the mixture shaken for 24 h at room temperature. The CPG was isolated by vacuum filtration, washed with dichloromethane and ether and dried *in vacuo* overnight. The amount of nucleoside loaded onto the support was determined by measuring the absorbance of the trityl cations (DMT^+) released from an accurately weighed amount of CPG upon treatment with 3% TCA in 1,2-dichloroethane (DCE). Prior to chain assembly on an oligonucleotide synthesizer, any free amino or hydroxyl groups present on the support were acetylated in order to "cap" any potential reactive sites.

Method B (Controlled and High Loading CPG). [236]

The rapid derivatization of controlled loadings of nucleoside (*e.g.* low nucleoside loadings: 5-10 μmol/g) and high loading supports (≈90 μmol/g) are attainable when the condensing reagents used are a mixture of either HATU or HBTU and 4-DMAP. When a specific loading of nucleoside was desired (*e.g.* 10 μmol/g), then a limiting amount (15-20 μmol) of nucleoside/gram (*i.e* 15-20 μmol/g) of succinylated CPG was used. When maximal loading were desired (*e.g.* 90 μmol/g) then an excessive amount of nucleoside/gram (400 μmol/g) of CPG was used. The conditions for the derivatization

of a high-loading CPG were as follows: succinylated LCAA-CPG (0.25 g), HATU or HBTU (0.1 mmol), 5'-O-DMT-N-protected nucleoside (0.1 mmol) and 4-DMAP (12 mg) were added to a septum-capped 10 mL glass vial. The coupling reaction was initiated by the addition of acetonitrile (1-2 mL) and the contents left shaking at room temperature for 2 hours. The CPG was filtered and washed successively with dichloromethane, methanol and ether and dried *in vacuo* overnight. The nucleoside loading was determined by trityl cation release from the CPG surface upon treatment with a known volume of 3% TCA in DCE. The absorbance reading of the trityl cation (DMT^+) was measured at 505 nm. Finally, the CPG was "capped" on the oligonucleotide synthesizer prior to chain assembly.

7.3.4 Solid-Phase Synthesis of Branched Oligonucleotides

7.3.4.1. Non-Regiospecific (Convergent) Synthesis[204,206,233]

Branched oligonucleotides (bDNA and bRNA) containing symmetrical (non-regiospecific) 2' and 3'-extensions surrounding the branchpoint nucleotide were conducted on a 1 μmol synthesis scale and typically utilized high loading CPG (50-90 μmol/g) for maximal branching efficiency.[233] Conventionally, synthesis was effected on 500 Å CPG, however the synthesis of long RNA molecules (>40 nucleotides) necessitated a larger pore diameter (1000 Å) for adequate coupling. The branching nucleoside, adenosine-2',3'-O-bis-phosphoramidite monomer (Compound **1.1**; section 7.8.5) was dissolved in anhydrous CH_3CN at a concentration of 0.03 M using a minimum of 100 mg of monomer. This ensured that (unavoidable) traces of moisture would not consume significant amount of the bis-phosphoramidite during the coupling (branching) step thus reducing the overall yield of bNA. Synthesis was conducted using the standard cycle (ABI 381A) supplied by the manufacturer with modifications listed above (section 7.3.2), and an extended coupling time of 30 minutes was used for the bis-phosphoramidite branching synthon. Oligonucleotides were cleaved from the support, and all base and phosphate protecting groups removed under standard conditions (section 7.4). In the case of RNA oligonucleotides a subsequent treatment with

desilylating reagent (TREAT-HF) was necessary in order to unblock the 2'-O-*tert*-butyldimethylsilyl protecting group. The amount of crude oligonucleotide was quantified by UV spectroscopy and analyzed by either denaturing PAGE (section 7.5.2) or anion exchange HPLC (7.5.3). Accordingly, the oligomers were purified by either preparative PAGE or HPLC, desalted by size-exclusion chromatography on Sephadex G-25®, requantified and concentrated to dryness (Speed-Vac® Concentrator, Savant Industries, NY).

7.3.4.2. Regiospecific (Divergent) Synthesis

Branched DNA oligonucleotides consisting of varied sequences at the 2' and 3' positions of the branchpoint were synthesized according the method of Braich and Damha.[221] Synthesis was conducted on a 1 μmol scale using low loading of CPG (5-10 μmol/g) in order to ensure ample extension from the branchpoint nucleotide. The 3',5'-linked linear portion was assembled using the standard synthesis cycle, 3'-DNA phosphoramidites and reagents, and the branching nucleotide introduced with an adenosine 3'-RNA phosphoramidite. In the case of Y-shaped (or forked) DNA molecules, chain elongation was continued in the 3'→5' direction followed by automated detritylation of 5'-position of the last nucleotide. The terminal 5'-hydroxyl was "capped" with acetic anhydride on the synthesizer, washed thoroughly with CH_3CN and dried with a stream of argon.

The column containing the support bound oligonucleotide was removed from the synthesizer and fitted with luer-lock syringe adapter. The phosphate β-cyanoethyl protecting groups were removed by pushing a solution of triethylamine/acetonitrile (4:6 v/v, 10 mL) through the column by syringe over a 90-minute period. This converted the phosphotriester moieties to the more stable phosphodiesters, which withstand the conditions required to remove the branchpoint silyl-protecting group in the ensuing step. Furthermore, the heterocyclic base protecting groups and CPG tether were unaffected by this treatment. The CPG beads were washed extensively with CH_3CN (30 mL) and THF (30 mL). The 2'-TBDMS group was selectively removed by washing the CPG with a solution of TBAF (1M in THF, 1 mL) for 10 minutes. Prolonged treatment with

TBAF should be avoided as this has been shown to cleave the oligonucleotide from the solid-support.[212] The support was washed sequentially with THF (50 mL) and CH_3CN (50 mL), reinstalled on the synthesizer and dried by flushing argon through the column for 15 minutes. The branch was synthesized by extending the chain from the 2'-hydroxyl of the riboadenosine unit with inverted 5'-DNA phosphoramidites. Owing to the steric hindrance surrounding the 2'-position, the first phosphoramidite addition necessitated a higher concentration (0.3 M) and coupling time (30 minutes).[219,220] All successive couplings utilized standard conditions. Complete deprotection, analysis and purification of oligonucleotides were identical to those for symmetrical bNAs (section 7.3.4.1).

7.4 Complete Deprotection of Synthetic Oligonucleotides

7.4.1. General Reagents

Sterilized water was prepared by treating double distilled and deionized water (Millipore; Billerica, MA) with 0.1 % (v/v) diethyl pyrocarbonate (DEPC; Aldrich) in a glass bottle, and stirred at room temperature for at least 2 hours thus ensuring that any RNases were inactivated by covalent acylation. Subsequently, it was autoclaved in an electric steam sterilizer at *ca.* 121°C (250 F, 15 PSI) for 1h to eliminate residual DEPC. All plasticware, glassware and pipet tips were sterilized by autoclaving for 1h at 121°C in the presence of DEPC. Following deprotection, gloves were worn at all times when handling the oligonucleotides in order to minimize contamination, and degradation from nucleases present on the skin. Water for HPLC analysis and purification of oligonucleotides was double distilled and deionized and pre-filtered through a 0.45 μM nylon membrane to remove any residual particulates and contaminants. Solutions and buffers for the analysis of oligonucleotides were made with DEPC-treated water, filtered through a 0.45 μM syringe filter unit (Millipore) and stored at -20°C. Ammonia (29%), absolute ethanol, *n*-butanol, and triethylamine trihydrofluoride (TREAT-HF) were used as received.

7.4.2. Cleavage from CPG and Deblocking of Phosphate, Base and Silyl Protecting Groups.

CPG-bound oligonucleotides were transferred to a 1.5 mL microtube and suspended in 3:1 (v/v) aqueous ammonia (29%)/absolute ethanol and gently shaken at r.t. for 24 h (48 h for those sequences containing N^2-(*i*-Bu)-guanine) to cleave the oligonucleotide from the support and deblock any phosphate and base protecting groups. The supernatant was removed, the CPG washed with ethanol (3 × 0.5 mL) and the fractions dried in a Speed-Vac® concentrator under a low (house) vacuum.

Given that RNA oligonucleotides bear an extra 2'-TBDMS protecting group, such molecules were consequently treated with the desilylating reagent, TREAT-HF (5 μL/crude A_{260} unit) for 48 h at r.t.[182] The solution was then either quenched with sterile water (1 mL) and dried, or precipitated directly from the desilylation reaction by adding 25 μL of 3 M sodium acetate (pH 5.5) followed by 1 mL of cold *n*-butanol.[294] The precipitated material was centrifuged at maximum speed for 10 min and the RNA pellet was washed with 70% ethanol (2 × 0.5 mL) and dried. Alternatively, the 2'-O-TBDMS group was removed using a mixture of TREAT-HF and N-methylpyrrolidinone (NMP; Aldrich) according to the method of Wincott *et al.*[293] Briefly, the crude, silylated RNA was suspended in 6:3:4 (v/v/v) NMP/triethylamine/TREAT-HF (250 μL), heated to 65°C for 1.5 h and precipitated with *n*-butanol as described above.

7.5. Analysis and Purification of Synthetic Oligonucleotides

7.5.1. General Reagents

High quality electrophoresis grade acrylamide, *N,N'*-methylene-bisacrylamide (BIS), ammonium persulfate (APS), *N,N,N',N'*-tetramethylethylenediamine (TEMED) and urea were obtained from Amersham. Bromophenol blue (BPB), xylene cyanol (XC), 2-Amino-2-(hydroxymethyl)-1,3-propanediol (Tris), boric acid, and ethylene-diamine

tetraacetate dihydrate (EDTA) were purchased from Bio-Rad (Mississauga, ON). Formamide and Stains-All® were obtained from Sigma.

Formamide was deionized by stirring over mixed bed ion-exchange resin (Bio-Rad AG 501-X8) for 30 minutes and filtered. All aqueous electrophoresis solutions were made with double distilled and deionized water (Millipore). Non-denaturing acrylamide gel solutions (12-24%) were prepared by mixing the appropriate volume of a 50% (w/v H_2O) acrylamide stock solution (19:1 acrylamide:BIS), 10 × TBE buffer (890 mM EDTA, 890 mM boric acid, 25 mM EDTA, pH 8.3) and diluting with water. In addition, acrylamide solutions for denaturing gels contained 7 M or 8.3 M urea. Sample loading buffer (denaturing) consisted of 8:2 (v/v) deionized formamide:10 × TBE. Running dye markers contained 2% (w/v) of both XC and BPB dissolved in sample loading buffer.

HPLC grade lithium perchlorate ($LiClO_4$) and *n*-propanol were purchased from Aldrich and Caledon (Georgetown, ON), respectively. Solutions of $LiClO_4$ for HPLC analysis and purification were filtered through a 0.45 μM nylon membrane and degassed under vacuum prior to use.

7.5.2. Analysis and Purification by Polyacrylamide Gel Electrophoresis (PAGE)

Fully deprotected oligonucleotides were predominantly analyzed and purified by denaturing PAGE using a vertical slab electrophoresis unit (Hoefer Scientific SE600; Fisher Scientific). Typically, analytical (0.75 mm) and preparative (1.5 mm) PAGE were run on 12-24% (w/v) acrylamide gels containing 1 × TBE running buffer (89 mM Tris/boric acid, 2.5 mM EDTA, pH 8.3). In general, oligonucleotide samples were run on 7 M urea gels, however, self-complementary oligonucleotides (*e.g.* dumbbells) were resolved on 8.3 M urea gels in order to prevent intramolecular or intermolecular association interactions. Gel solutions (30 mL) were degassed by sonication (10 minutes), and polymerization was initiated by the addition of 10% (w/v H_2O) ammonium persulfate (APS; 200 μL) and TEMED (25 μL). Prior to sample loading, the

individual wells were flushed with 1× TBE running buffer in order to remove any residual debris and urea. Crude samples were dissolved in 15 μL or 50-100 μL of loading buffer for analytical and preparative gels respectively, and heated at 90°C for 5 minutes. Gels were run at 500 V for 30 minutes followed by 800 V until the faster moving BPB dye was 2 inches from the bottom of the gel (*ca.* 2.5 h).

Following electrophoresis, the gel was carefully wrapped in plastic film (*e.g.* Saran Wrap), placed over a fluorescent TLC plate, and the resolved oligonucleotide band(s) were illuminated using a handheld UV lamp (λ=254 nm). The UV-illuminated gels were photographically captured using a Polaroid PolaPan® camera and instant film (Polaroid #52, 4"×5", ISO 400/21°C) through a Kodak Wratten gelatin filter (#58 green). Equally, the illuminated gels images were digitally captured on a Kodak DC3800 camera. Alternatively, analytical gels were visualized by soaking in a solution containing Stains-All® (Sigma), which was prepared according to manufacturer's specifications, and photographed without UV illumination.

For preparative gels, the desired oligonucleotide band was excised from the gel with a scalpel blade, placed in a sterile culture tube (15 mL), crushed to fine particles and soaked in autoclaved water (5 mL) for 24-48 h.[453] Similarly, the crushed gel pieces were suspended in 0.1 M sodium acetate buffer, pH 6 (3 mL), heated at 90°C for 5 minutes and rapidly frozen at -70°C for 5 minutes.[454] The contents were rapidly thawed at 90°C. In both methods, the slurry was centrifuged to settle much of the gel debris, and the supernatant dried. The purified oligonucleotides contained high levels of low molecular weight impurities such as salts and urea, and were consequently "desalted" by size exclusion chromatography on Sephadex G-25® (section 7.5.5.1) or reverse-phase Sep-Pak® cartridges (section 7.5.5.2).

7.5.3. Analysis and Purification of Synthetic Oligonucleotides by Anion-Exchange HPLC

Linear and branched oligonucleotides may be easily and efficiently analyzed and purified from the crude mixture by anion exchange HPLC. Such analyses were conducted on a Waters 480 HPLC instrument equipped with dual 501 pumps, a UK6 injector, internal column heater and a 480 tunable UV detector. The mobile-phase mixture was controlled by a 600E gradient controller and solvent delivery system. Chromatograms were accumulated and processed with Millenium-32 (version 2.0) software. Alternatively, a Waters Breeze® system consisting of a 1525 binary pump, a 2487 dual absorbance detector, a Rheodyne® manual injector and internal column heater was used. Analyses and preparatory injections were conducted on a Waters Protein-Pak™ DEAE 5PW (7.5 mm × 75 mm) column with a variable linear gradient of: deionized water (Buffer A) and 1 M $LiClO_4$ (Buffer B) and over the specified time, according to the size of the oligonucleotide to be analyzed (**Table 7.1**). The column was heated to 50-60°C in order to denature any intramolecular or intermolecular interactions between self-complementary oligonucleotides.

Table 7.1: Conditions for the analysis and purification of linear and branched oligonucleotides by anion-exchange HPLC.[a]

Length of Oligonucleotide (nt)	Buffer A (%)[b]		Buffer B (%)[c]		Gradient Run Time (min)
	Initial	Final	Initial	Final	
<5	100	90	0	10	60
6-15	100	80	0	20	60
>15	90	80	10	20	60

Waters Protein Pak DEAE-5PW Column (7.5 mm × 75 mm); column temperature = 50°C. [b]Buffer A = deionized water; [c]Buffer B = 1 M lithium perchlorate or sodium perchlorate. nt=nucleotides.

Prior to analysis, samples were heated to 60°C for 5 minutes, and centrifuged at 14 Krpm for 10-15 seconds in order to settle any particulates that may clog the injection loop. Analytical injections typically contained 0.1-0.5 A_{260} units of oligonucleotide dissolved in 25-100 μL of water, whereas preparatory injections had 30-60 A_{260} units dissolved in 250-500 μL of water. Higher loadings (>60 A_{260} units) were shown to

overload the column and compromise separation of the desired oligonucleotide from failure sequences present in the mixture.

In the case of preparatory (*i.e.* purification) runs, the absorbance was monitored at 290 nm to avoid saturation of the detector signal. Fractions corresponding to the peak of interest were collected in sterile 1.5 mL microtubes (1 mL fractions). The anticipated retention time was similar to that obtained during routine HPLC analysis of the same oligonucleotide. The fractions were dried down, and pooled in 0.25 mL of sterile water. The oligomers were precipitated from the perchlorate salts using 4 volumes (1 mL) of cold *n*-propanol.[294] Small oligonucleotides (≤5-mers) do not precipitate out of solution efficiently. In this case, they were desalted by size exclusion chromatography on Sephadex G-25® columns (section 7.5.5.1) or by reverse phase chromatography on Sep-Pak® cartridges (section 7.5.5.2). Lithium perchlorate is much more soluble in organic solvents than other perchlorate salts, making precipitation easy and efficient, thus preventing a final desalting step for larger oligomers. The samples were cooled at -20°C for 4-6 hours, and centrifuged at 14 Krpm for 10 minutes. The supernatant was discarded and the white pellet washed with cold *n*-propanol (2 × 0.25 mL). The pellet was dried in a Speed-Vac® concentrator, and the pure oligonucleotide quantified by UV spectroscopy.

7.5.4. Reverse-Phase HPLC

In some instances, oligonucleotide samples were purified by reverse-phase HPLC. This method is advantageous as is precludes the need for a final precipitation or desalting step given that the mobile phases used are completely volatile, and can be removed by evaporation under reduced pressure. The mobile phases used were; Buffer A: 0.1 M triethyl ammonium acetate (TEAA, pH 6.5) and Buffer B: acetonitrile (HPLC grade, Caledon). The column was heated to 60°C for the duration of the run. Briefly, the oligonucleotide samples were dissolved in either sterile water or buffer (0.1 M TEAA pH 6.5) and injected onto a Waters Symmetry® C-18 reverse-phase column (4.6 × 150 mm, 5 μM particle size) using a linear gradient in Buffer B consisting of: 100% Buffer

A for 5 minutes, then 95% Buffer A from 5-10 minutes followed by 5-20% Buffer A from 10-60 minutes. Typically, the oligonucleotide samples eluted between 25-30 minutes under these conditions. The peaks of interest were collected in 1.5 mL microtubes, dried in a Speed-Vac® concentrator and the pure oligonucleotide quantified by UV spectroscopy.

7.5.5 Desalting of Oligonucleotides.

7.5.5.1. Size Exclusion Chromatography on Sephadex G-25®

Purified oligonucleotides were isolated from any low molecular weight organic species and water soluble salts by gel filtration on a highly crosslinked oligosaccharide (*i.e.* dextran) matrix. Macromolecules (*e.g.* oligonucleotides) are completely excluded from the bead pores, whereas smaller molecules (*e.g.* salts) penetrate the beads to varying extents depending on the size of the molecule. Sephadex G-25® was purchased in its dry form from Amersham, and swollen in water for 2 hours prior to autoclaving in the presence of DEPC (0.1 % v/v) at 121°C for 1 hour. The sephadex (10 mL) was poured into a sterile plastic syringe barrel (10 mL) plugged with silanized glass wool, and washed with 3 column volumes of DEPC water. The "salted" sample, dissolved in 1 mL of water, was loaded onto the column and eluted with water (10 mL). The eluted oligonucleotide was collected in 1 mL fractions and the amount of oligonucleotide quantified by UV spectroscopy at 260 nm. Typically the pure, desalted oligonucleotide was in the first four fractions. Customarily, the recovery of material after desalting was between 70-90%. The fractions containing the oligomers were pooled, dried, redissolved in sterile water (1 mL) and this stock solution stored at -20°C.

7.5.5.2. Reverse-Phase Chromatography of Sep-Pak® Cartridges.[203]

In the rare instance that samples could not be desalted by size-exclusion chromatography on Sephadex-G25® (*e.g.* small oligomers), oligonucleotides were alternatively desalted by reverse-phase chromatography after PAGE or HPLC purification. Sep-Pak® cartridges (Waters; Mississauga, ON) employ the principals of reverse-phase liquid chromatography to isolate sample components. The cartridge was attached to a 10 mL

plastic syringe and flushed consecutively with HPLC grade methanol (10 mL) and water (10 mL). The "salted" oligonucleotide sample was dissolved in 50 mM triethylammonium acetate buffer (TEAA; 1 mL), loaded onto cartridge and the eluent collected dropwise. It was subsequently flushed with 50 mM TEAA (3 mL) and the oligonucleotide eluted with 7:3 50 mM TEAA/methanol (10 mL) collecting 1 mL fractions. The fractions containing the oligonucleotide, typically the first 2-3 fractions, were combined, dried and quantified by UV spectroscopy. The pure oligonucleotide was redissolved in sterile water (1 mL) and stored at -20°C.

7.5.6 Characterization of Oligonucleotides by MALDI-TOF Mass Spectrometry[455,456]

Matrix assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) was conducted on a Kratos Kompact-III instrument (Kratos Analytical Inc, New York) in either negative reflector mode (<30 nucleotides) or linear mode (>30 nucleotides). Samples were desalted by Sephadex-G25® or Sep-Pak® cartridges prior to analysis in order to minimize the adduction of cations, which shift the ion signal to higher *m/z* values, thus interfering with accurate determination of molecular weights. When a sample is ionized, normally singly charged ions of oligonucleotides are observed, which makes MALDI-MS spectra very easy to interpret. Matrices and co-matrices such as: 6-aza-2-thiothymine (ATT), spermine and fucose were of the highest purity and purchased from Sigma-Aldrich. All analytes (*i.e.* oligonucleotides) and matrices were diluted with ultrapure deionized water and HPLC grade solvents. Purified oligonucleotide samples (1 μL, 0.2 – 1 nmol/μL in water) were mixed with 1 μL of saturated ATT/spermine (80 mg/mL ATT dissolved in 1:1 water/CH_3CN containing 25 mM spermine) and 1 μL fucose (50 mM in water). The analyte/matrix mixture (1 μL) was spotted in duplicate on a sample plate and allowed to dry under a warm stream of air. The crystalline spots were analyzed and the spectra processed using software supplied by Kratos.

7.6. BIOPHYSICAL CHARACTERIZATION OF OLIGONUCLEOTIDES

7.6.1 Hybridization Studies: UV-Thermal Denaturation Studies

Thermal denaturation profiles (melting curves) were acquired on a Varian CARY 1 UV-Vis spectrophotometer (Varian; Mulgrave, Australia) equipped with a multiple cell holder, a Peltier thermal cell holder and temperature controller. Spectra were processed using Cary Win UV software (version 2.00). The hybridization properties of oligonucleotides were investigated by monitoring the change in UV-absorbance (λ=260 nm) with increasing temperature. Hybridization buffers consisted of either: (a) 10 mM Tris-HCl, 10 mM NaCl, pH 7.5 or (b) 0.25 M MES, 20 mM $MgCl_2$ pH 7.6 or (c) 10 mM Na_2HPO_4, 0.1 mM Na_2EDTA, pH 7.0. Complementary stands were dissolved in fixed ratios in 0.5-1 mL of buffer at a concentration of 4-10 μM of oligonucleotide single strands. Oligonucleotide extinction coefficients (ε_{260}) were calculated by applying the nearest-neighbor approximation of Puglisi and Tinoco[14] using an internet-based biopolymer calculator. Oligonucleotide mixtures were heated to 90°C for 10-15 minutes in order to dissociate any non-specifically bound regions, cooled slowly to room temperature for 30 minutes and then left at 4°C overnight. The annealed samples were transferred to pre-chilled Hellma QS-1.000 (Cat # 114) quartz cells, sealed with a Teflon-wrapped stopper and degassed by sonication for 15 seconds. The complexed oligonucleotides were equilibrated to 5°C in the cell holder of the spectrophotometer for 5 minutes prior to spectral acquisition. The absorbance at 260 nm was measured at 0.5°C intervals at a temperature rate of 0.5 °C/minute. The thermal melting temperature (T_m) values were calculated as the maximum of the first derivative plots of the absorbance *versus* temperature profiles, and coincides with the point at which half of the complexed oligonucleotides are in their single-stranded state. Spectra were typically acquired in duplicate (or triplicate) and the calculated T_m's were consistently within 0.5-1°C of each other. The data obtained was transferred to spreadsheet software (Miscrosoft® Excel 97) for subsequent analysis and presentation. Comparative hyperchromicity values (*i.e.* changes in relative absorbance) were obtained by using the fomula: $H=(A_T-A_0)/A_f$, where H is the hyperchromicity, A_T is the absorbance at any

given temperature (T), A_0 is the initial absorbance reading, and A_f is the absorbance at the highest temperature.[14] Alternatively, normalized absorbance values (between 0 and 1) were calculated in order to compare plots of unequal hyperchromicity, such as those containing non-complementary regions (*e.g.* lariat DNA dumbbell) according to the equation: $A_{norm}=(A_t-A_0)/(A_f-A_0)$.[457]

7.6.2 Circular Dichroism Spectroscopy (CD)

Circular dichroic spectra were accumulated on a JASCO J-710 spectropolarimeter. Samples were contained in a Hellma QS-1.000 (Cat # 114) fused quartz cell, which was cooled by an external circulating bath (VWR Scientific) at constant temperature (5°C). The data was acquired using a personal computer interfaced to the spectrometer processor. The spectra were representative of 5 individual scans recorded between 220-340 nm. Individual scans were at a rate of 50 nm/minute using a sampling wavelength of 0.2 nm and a bandwidth of 1 nm. Oligonucleotide samples were dissolved in 0.5-1 mL of buffer at a concentration of 4-10 μM of oligonucleotide single strands (Buffer: 10 mM Tris-HCl, 10 mM NaCl, pH 7.5). Prior to analysis, samples were preheated and cooled as for T_m analysis (section 7.6.1), left at 4°C overnight and allowed to equilibrate in the instrument's cell holder to the appropriate sampling temperature for at least 10 minutes. The acquired data was processed using the J-700 Windows software (Version 1.00) supplied by the manufacturer (Jasco Inc.). The CD spectra were background subtracted (*i.e.* buffer subtraction), smoothed and were corrected for concentration such that the molar ellipticity could be determined. The corrected data was exported to Microsoft® Excel spreadsheet software for further analysis and presentation.

7.7. General Molecular Biology Techniques: oligonucleotide Labeling & characterization

Synthetic oligonucleotides used for molecular biology purposes were purified either by preparative PAGE or anion exchange HPLC, and desalted by size exclusion chromatography on Sephadex G-25® or Sep-Pak® cartridges.

7.7.1. 5'-End labeling of Synthetic Oligonucleotides

Synthetic branched and linear oligonucleotide substrates (DNA and RNA) were labeled at the 5'-hydroxyl terminus with a radioactive phosphorus probe using the enzyme T4 polynucleotide kinase (T4 PNK) according to manufacturer's specifications (Pharmacia). Labeling was performed in 50 μL reaction volumes consisting of: oligonucleotide substrate (10 pmol), 5 μL of T4 PNK 10 × reaction buffer (0.5 M Tris-HCl, pH 7.6, 100 mM $MgCl_2$, and 100 mM 2-mercaptoethanol), 18 U of T4 PNK enzyme (diluted to 3 U/μL in a solution of 50 mM Tris-HCl, pH 7.5, 1 mM, 10 mM 2-mercaptoethanol, 0.1 mM EDTA and 50% glycerol), 20 pmol of [γ-^{32}P]-ATP (6000 Ci/mmol, 10 mCi/mL; Pharmacia) and the remaining volume made up with sterile water.

The reaction mixture was incubated at 37°C for 0.5-1 h at which time the enzyme was deactivated by heating at 65°C for 10 minutes. The reactions were dried down in a Speed-Vac® concentrator, and dissolved in 2.5 μL each of sterile water and gel loading dye (98% v/v deionized formamide in 10 × TBE, 1 mg/mL BPB, 1 mg/mL XC and 10 mM EDTA) and loaded onto a denaturing polyacrylamide gel (16%, 7M urea). The "hot" gel was peeled from the glass surface using a previously exposed autoradiogram film and wrapped with Saran Wrap®. The labeled oligonucleotides were visualized by autoradiography (0.5-1 min, room temperature) in a Kodak X-Omatic® cassette with an intensifying screen on X-OMAT-AR film (Kodak). The autoradiogram was laid over the wrapped gel, the desired bands circled with a permanent marker on the gel and the most intense oligonucleotide band excised from the gel. The gel piece was crushed in a

1.5 mL microtube, extracted into 0.5 mL of sterile water at 37°C overnight, desalted on NAP-10® (size exclusion) columns (Amersham), and stored at –20°C.

7.7.2. 3'-End Labeling of DNA Oligonucleotide Substrates with Terminal Deoxynucleotidyl Transferase (TdT)

Linear and branched DNA molecules were radioactively labeled with 5'-[α-^{32}P]-dideoxyadenosine monophosphate at their free 3'-hydroxyl termini using the enzyme terminal deoxynucleotidyl transferase (TdT; Amersham).[318] Reactions were conducted in 50 μL total volumes consisting of: DNA oligonucleotide substrate (10 pmol), 10 μL of 5× terminal deoxynucleotidyl transferase buffer (0.5 M sodium cocadylate, pH 7.2, 10 mM $CoCl_2$, 1 mM 2-mercaptoethanol; Amersham), 17-20 pmol of [α-^{32}P]-ddATP (3000 Ci/mmol, 10 mCi/mL; Amersham), 5 μL of terminal deoxynucleotidyl transferase enzyme (11 U/ μL) and water to 50 μL total volume. Incubation at 37°C for 2 hours proceeded followed by deactivation of the enzyme at 65°C for 10 min. Subsequently, the mixtures were dried down in a Speed-Vac® concentrator, and dissolved in 2.5 μL each of sterile water and gel loading dye, loaded onto a denaturing polyacrylamide gel (16%, 7M urea), autoradiographed and purified as described previously for the 5'-end labeled substrates.

7.7.3. Characterization of Synthetic Lariats by Debranching with HeLa Nuclear Extract[109]

Radioactively labeled (5'-end or 3'-end) or unlabelled branched and lariat substrates (1-200 pmol) were dissolved in 7 μL of buffer comprised of 20 mM HEPES, pH 7.6, 100 mM KCl, 20% glycerol and 0.5 mM DTT and also containing 10 mM EDTA as a general 3'-exonuclease inhibitor.[109] HeLa nuclear extract (3 μL) was added and the reactions incubated at 30°C for 0.5-1 h. The samples were either phenol/chloroform extracted prior to gel loading or directly loaded onto a denaturing gel (12-24%) which had been pre-run at 500 V for 15 min. Radioactively labeled substrates were visualized

by autoradiography and non-radioactively labeled substrates were visualized by staining the gel for 16-24 h in Stains-All® solution (Sigma Aldrich) prepared according to manufacturer's protocols.

7.7.4. Characterization of Lariats by Digestion with Bovine Spleen Phosphodiesterase (BSPDE)

Bovine spleen phosphodiesterase (BSPDE) was obtained as a lyophilized powder (10 U/mg; Sigma-Aldrich) and diluted with sterile water to an activity of 0.1 U/μL. Oliogonucleotides radiolabelled at the 3'-termini (100 fmol-1 pmol) were dissolved in 48 μL of 0.1 M sodium acetate, pH 6.5 and 2 μL of BSPDE (0.2 U) added. The reactions were incubated at 37°C for 0.5-1 hour, and dried. The samples were redissolved in 5 μL of sterile water and 5 μL of gel loading dye, run on a 12-16% denaturing gel (7 M urea) and visualized by autoradiography.

7.8. CHAPTER 2: SOLID-PHASE SYNTHESIS OF LARIAT DNA ON MIXED LINKER (Q/S-LINKER) SOLID-SUPPORT

Our postulated methodology for the synthesis of lariat-oligonucleotides on solid-support takes advantage of our recognized methods for the convergent synthesis of branched nucleic acids (Y-shaped), and the differential rate of cleavage of two CPG tethers, the succinyl-linker (*S-linker*) and the hydroquinone-O,O'-diacetyl linker (*Q-linker*). Prior to selective cleavage of the *Q-linker*, the 5'-hydroxyl terminal of the molecule is phosphitylated and then after Q-cleavage, subsequently activated using standard methods in order to yield a new phosphodiester linkage with the concomitant cyclization of the 5' and 3' termini.

7.8.1. Model Synthesis of a Thymidine Dimer on Solid Support

In order to assess if the phosphitylation of a terminal 5'-hydroxyl group followed by activation with tetrazole could result in the formation of a new phosphodiester linkage, a model system was designed to test the efficiency of coupling of a 5'-phosphitylated CPG-bound thymidine residue with the free 3'-hydroxyl of a dT nucleoside. 5'-DMT-2'-deoxythymidine-3'-succinyl-LCAA-CPG (1 μmol) was detriltylated using 3% TCA in DCE on a DNA synthesizer using a standard detritylation cycle (ABI 381A). The solid-support was rinsed thoroughly with CH_3CN (10 mL) and dried by reverse flush under a stream of argon for 10 min. The column was then placed under a dessicated vacuum environment to ensure that the CPG was absolutely dry. The CPG was suspended in CH_3CN (250 μL) to which was added, anhydrous pyridine (2 μL, 0.1 M) and the phosphorylating reagent, β-cyanoethyl-(N,N-diisopropylamino) phosphorochloridite (1.1 μL, 0.04 M, 10 eq.). The reaction was gently shaken at r.t. for 45 min-1h, and filtered through a fritted micro-spin column. The support was washed consecutively with CH_3CN (10 × 0.2 mL) and diethyl ether (2 × 0.2 mL) and placed under vacuum to dry.

Nucleoside Coupling to the 5'-Phosphoramidite. The dried, phosphitylated CPG was treated with 5'-DMT-2'-deoxythymidine (100 μL, 0.1 M in CH_3CN) and tetrazole (100 μL, 0.5 M in CH_3CN) and gently shaken at r.t for 45 min. Oxidation of the resultant phosphite triester to the phosphotriester was conducted *in situ* by adding a solution of 0.1 M iodine (200 μL; dissolved in 25:20:2, THF/pyridine/H_2O) and left shaking for an additional 5 minutes. The contents were filtered through a micro-spin column and washed successively with CH_3CN (10 × 0.2 mL) and diethyl ether (2 × 0.2 mL) and dried. The support bound dimer was treated with detritlyating reagent (3% TCA) on the synthesizer and the DMT cation quantified. Comparison of the trityl absorbances (at 505 nm) of the first CPG-bound nucleoside and the dimer indicated an overall coupling of *ca.* 65%. In order to verify the true extent of coupling, the support-bound dimer was treated with 3:1 NH_4OH/ethanol for 1 hour at r.t in order to cleave the dimer from the support and remove the β-cyanoethyl protecting group. Formation of the thymidine dimer was monitored by PAGE (24%, 7M urea) and anion exchange HPLC (0-10% linear gradient in 1M $LiClO_4$) and compared with authentic standards of thymidine dimer, thymidine-5'-monophosphate and thymidine. Integration of the resultant HPLC peaks revealed that the dimer was formed in 71% yield. The unreacted thymidine nucleoside was present in 15% yield, whereas the uncoupled thymidine-5'-monophosphate existed in 14% yield.

7.8.2. Monomer Synthesis: 5'-O-dimethoxytrityl-thymidine-3'-O-hydroquinone-O,O'-diacetyl hemiester (5'-DMT-dT-Q; Compound 2.1)[167,168]

The 3'-hydroquinone hemiester functionalized thymidine (5'-DMT-dT-Q) nucleoside was prepared by combining: 5'-O-dimethoxytrityl thymidine (0.5 g, 0.92 mmol), hydroquinone-O,O'-diacetic acid (0.29 g, 1.3 mmol; Lancaster Synthesis Ltd., Lancashire, UK), DEC (176 mg, 0.92 mmol) and 4-DMAP (24 mg, 0.2 mmol). The contents were dissolved in pyridine (10 mL) and triethylamine (0.1 mL) in an oven-dried flask. The reaction was left stirring at r.t. for 16 h, and monitored by TLC (9:1, chloroform:ethanol). Unreacted nucleoside was consumed by adding additional DEC (0.3 mmol) and the reaction stirred for a further 8 h. The pyridine solution was

concentrated to an oil, redissolved in chloroform (50 mL) and washed with water (3 × 50 mL). The organic layer was dried over $MgSO_4$, filtered and evaporated *in vacuo* to yield a brown oil consisting of the predominant hemiester (*ca.* 75%) and some of the dinucleoside diester (R_f= 0.53).

The crude material was redissolved in a minimal amount of chloroform, applied to a silica-gel column and eluted sequentially with chloroform (500 mL), 5% ethanol/chloroform (500 mL) and 30% ethanol/chloroform (250 mL) to yield the pure product in 62% yield. TLC: R_f (9:1 chloroform/ethanol)=0.06, FAB-MS (NBA matrix): $[M]_{calculated}$=752.8, $[M-H]^-_{observed}$=751.3.

7.8.3. Monomer Synthesis: 5'-O-dimethoxytrityl-thymidine-3'-O-succinic acid hemiester (5'-DMT-dT-S; Compound 2.2)

The 3'-succinate of the thymidine nucleoside (5'-DMT-dT-S) was prepared according to published methods with slight modifications.[167,235], or was purchased from a commercial supplier (ChemGenes Corp.).

The protected nucleoside, 5'-DMT-thymidine (0.5g, 0.92 mmol) was combined in an oven-dried flask with succinic anhydride (110 mg, 1.1 mmol), DEC (176 mg, 0.92 mmol) and 4-DMAP (56 mg, 0.46 mmol). The contents were dissolved in pyridine (10 mL) and left stirring for 24 h (r.t.). The pyridine was evaporated *in vacuo* and the residual oil dissolved in chloroform (50 mL). The solution was washed with saturated salt solution (3 × 50 mL), dried over Na_2SO_4, filtered and evaporated to yield the crude product as a pale brown foam.

The crude thymidine succinate was dissolved in 1% ethanol/chloroform and purified by silica-gel chromatography employing a gradient of 1-10% ethanol in chloroform to yield the pure product in 51% yield. TLC: R_f (9:1 chloroform/ethanol)=0.24, FAB-MS (NBA matrix): $[M]_{calculated}$=644.7, $[M+Na]^-_{observed}$=667.0.

7.8.4. Derivatization of Mixed-Q/S-Linked CPG

The derivatization of CPG with a mixture of 5'-DMT-dT-Q and 5'-DMT-dT-S nucleosides involves the coupling of equimolar amounts of the two different protected nucleosides to acid-activated LCAA-CPG using the uronium salts HATU or HBTU and 4-DMAP. For instance, when a loading of 70 μmol/g of the mixed Q- and S-linked nucleosides (*i.e.* ≈35 μmol/g Q-linked and ≈35 μmol/g S-linked) was desired, the reaction typically consisted of: acid-activated LCAA-CPG (0.2 g), 5'-DMT-dT-Q nucleoside (14 μmol, 11 mg), 5'-DMT-dT-S nucleoside (14 μmol, 9 mg), HATU (50 μmol, 19 mg) and 4-DMAP (30 μmol, 4 mg). The contents were dissolved in acetonitrile (2 mL) and the mixture shaken 2 h (r.t.). The CPG was filtered and washed thoroughly with methanol and chloroform and dried *in vacuo* for several hours.

Determination of the Q + S-linker Loadings. Cleavage of all the Q-linked nucleosides from the support proceeded by treating a small amount of mixed-Q/S-CPG (15 mg) with ammonium hydroxide (29%; 2 mL) at room temperature for 5 minutes. The reaction was stopped by filtering off the liquid and washing the CPG with methanol and chloroform. The sample was dried and the S-linked nucleoside loading determined by measuring the trityl cation released from the CPG surface upon treatment with 3% TCA. The loading of S-linked nucleoside was determined to be 37 μmol/g. The amount of Q-linked nucleoside was determined by measuring the loading of the mixed Q/S-linked support and subtracting the loading of the S-linked support from this number.

$$Loading\ (Q) = Loading\ (Q + S) - Loading\ (S)$$

The loading of the Q-linked nucleosides was determined to be 33 μmol/g, with a total Q +S loading of 70 μmol/g. Supports with lower or higher Q+S-linked nucleoside loadings were obtained by adjusting the amount of nucleoside hemiesters added to the reactions (assuming a 50% coupling reaction), however always maintaining an equimolar mixture between the two.

7.8.5. Bis-Adenosine Phosphoramidite Monomer Synthesis (Compound 1.1): 5'-O-dimethoxytrityl-(N^6-benzoyl adenosine)-2', 3'-bis-O-(β-cyanoethyl) N,N'-diisopropyl phosphoramidite

The adenosine bis-phosphoramidite branching synthons (Compound **1.1**) for the convergent synthesis of bNAs was prepared as described previously.[187] The starting reagent, N^6-benzoyl-5'-O-(dimethoxytrityl)adenosine was either synthesized from adenosine using the transient benzoylation procedure of Ti *et al.*[171], followed by dimethoxytritylation of the 5'-hydroxyl group[238], or purchase from ChemGenes.

Briefly, N^6-benzoyl-5'-O-(4,4'-dimethoxytrityl)adenosine (1.30 g, 2.0 mmol), 4-DMAP (84 mg, 0.7 mmol), dry THF (6.0 mL) and DIPEA (3.6 mL, 21 mmol) were added to a dry round bottom flask. The mixture was stirred under an inert atmosphere until all of the starting material dissolved, upon which an excess of β-cyanoethyl N,N-diisopropylphosphoramidochloridite (1.8 mL, 8.3 mmol) was slowly added *via* a dry syringe. After a few minutes, the white diisopropylethylammonium hydrochloride salt was clearly visible, indicating effective progression of the reaction. The mixture was stirred at room temperature for 1 h or until all of the starting material was consumed. Monitoring of the reaction by TLC (1:1 dichloromethane:diethyl ether) revealed the appearance of a fast moving dumbell-shaped spot, indicative of the presence of a diastereomeric mixture of bis-phosphoramidite products, which are only partially resolved by the solvent system.

The reaction was taken up in 500 mL of ethyl acetate (pre-washed 3x with 5% (w/v) $NaHCO_3$) and washed with brine solution (5 x 100 mL). The organic layer was dried over Na_2SO_4, isolated by filtration and the solvent removed *in vacuo*.

Purification of the bis-phosphoramidite product proceeded by silica gel column chromatography using an isocratic solvent mixture of dicholoromethane, hexanes and triethylamine [50:47:3 (v/v/v)] to afford the pure bis-phosphoramidite in 52% yield. The product bis-phosphoramidite was confirmed by comparison with an authentic sample

(Dr. S. Robidoux). TLC: R_f (1:1 dichloromethane/diethyl ether)=0.51 & 0.40, FAB-MS (*p*-nitrobenzyl alcohol matrix): $[M]_{calculated}$=1074.17, $[M]_{observed}$=1074.57, UV λ_{max} (95% EtOH)= 279 nm, ^{31}P-NMR (400 MHz): diastereomer 1, 152.3 & 150.7 ppm ($^5J_{P\text{-}P}$=10.1 Hz); diastereomer 2, 151.9 & 150.5 ppm ($^5J_{P\text{-}P}$=6.6 Hz); diastereomer 3, 151.8 & 151.0 ppm ($^5J_{P\text{-}P}$=4.6 Hz); diastereomer 4, 151.3 & 150.2 ppm ($^5J_{P\text{-}P}$=8.9 Hz)

7.8.6. Synthesis of Branched-DNA on Q/S-linked CPG (Compound 2.4)

Bis-functionalized CPG (Q/S-linked) of various loadings (20-83 μmol/g) were derivatized as described in section 7.7.4. The synthesis of the branched 21-mer oligonucleotide, 5'-$T_{10}A^{2'}$5'$(T_{10})_{3',5'}T_{10}$ was conducted on a 1 μmol scale using the convergent methodology for the synthesis of branched oligonucleotides (bNAs) described in section 7.3.4.1. The synthesis was conducted in the "trityl off" mode in order to expose the free 5'-hydroxyl of the oligomers for the succeeding reaction. A small amount of CPG (*ca.* 0.25 μmol) was removed from the column and deprotected under standard conditions so as to evaluate the extent of branching in each reaction (*i.e.* 20% PAGE (7 M urea) and HPLC). The remaining CPG was dried under vacuum.

7.8.7. Phosphitylation of the 5'-Hydroxyl on Support

The residual CPG was placed in a septum-sealed, nitrogen-purged vial and suspended in CH_3CN (0.5 mL). Anhydrous pyridine (10 μL, 0.25 M) and β-cyanoethyl-(N,N-diisopropylamino) phosphorochloridite (10 μL, 0.1 M, 50 eq) were added and the mixture shaken gently at r.t. for 2 h. Unreacted reagents were washed off the CPG surface with CH_3CN (10 × 250 μL) and diethyl ether (2 × 250 μL) and left to dry under vacuum overnight.

7.8.8. Selective Hydrolysis of the Q-Linker[167]

The phosphitylated CPG (0.5 μmol) was treated with 0.05 M K_2CO_3 in dry methanol (0.3 mL) for 1 minute. The methanolic supernatant was isolated, concentrated to

dryness and redissolved in 10 mM $MgCl_2$ (100 μL). The oligonucleotides were precipitated with cold ethanol (95%, 0.5 mL) and placed at -20°C for 2 h. The mixture was centrifuged at 14 Krpm and the precipitated pellet washed with cold 95% ethanol (2 × 0.25 mL). The pellet was redissolved in sterile water (1 mL) and quantified to determine the amount of hydrolyzed Q,Q-linked oligonucleotides cleaved from the support. The intactness of the oligonucleotides was verified by PAGE analysis (20%, 7M urea). The remaining CPG was washed consecutively with CH_3CN (10 × 250 μL) and diethyl ether (2 × 250 μL) and dried.

7.8.9. Cyclization of Branched DNA on Solid-Support (Compound 2.5)

The hydrolyzed Q-linked CPG (containing only S-linked material) was placed in a nitrogen-purged glass vial. A solution of 0.5 M tetrazole in CH_3CN (0.4 mL) was added and shaken at r.t. for 16 hours. Oxidant solution (0.1 M I_2 in 25:20:2, THF/pyridine/water; 0.4 mL) was added, shaken for 5 minutes, filtered and washed with CH_3CN (10 × 250 μL) and diethyl ether (2 × 250 μL) and dried.

Oligonucleotides were completely deblocked under standard conditions (section 7.4) and quantitated by UV spectrophotometry at 260 nm. Analysis of the reaction mixture by 12-28% denaturing PAGE (7M urea) and anion exchange HPLC revealed the presence of several slower moving oligonucleotide products. Selective bands were isolated from PAGE, extracted into water, and desalted on Sephadex G-25®. Subsequently, the product bands were characterized by methods such as 5'-end labeling with T4 polynucleotide kinase (section 7.7.1), enzymatic digestion with a 3'-exonuclease, bovine spleen phosphodiesterase (BSPDE; section 7.7.4), debranching with HeLa extracts containing a selective 2'-phosphodiesterase (section 7.7.3), and MALDI-TOF mass spectrometry (section 7.5.6).

7.8.10. Convergent Synthesis of Dendritic DNA

A small amount of the CPG-bound branched oligonucleotide, $T_{10}A^{2'}$-5'-$(T_{10})_{3',5'}$-T_{10} (0.25 μmol, 90 μmol/g CPG loading) was placed into a clean synthesizer column, and the 5'-termini branched further using the bis-adenosine phosphoramidite reagent. The resultant "hyper-branched" product was completely deprotected under standard conditions, and the crude mixture used as a control during PAGE analysis of the cyclized products above.

7.9. CHAPTER 3: SYNTHESIS OF LARIAT-DNA *VIA* THE CHEMICAL LIGATION OF A DUMBBELL COMPLEX

A complete list of the sequences synthesized for this study along with their crude and purified isolated yields and characterization data is presented in **Table 7.2.**

Table 7.2: Crude yields, isolated yields and MALDI characterization of oligonucleotides used in the intramolecular (dumbbell) synthesis of lariat-DNA study.

Code	Sequence (5'→ 3')	Yield[a] (ODU)			MW (g/mol)	
		Crude	Pure		Calc.	Found
Bimolecular Complexes						
3.1	^{HO}cgc-$A_{3',5'}$-cgc$_{OH}$	40.3	5.6[b]	N	2082.4	2081.3
3.2	^{HO}cgc-$A^{2',5'}$-cgc$_{OH}$	43.9	3.4[b]	N	2076.4	2081.3
3.3	gcgtgcg	47.6	15.9	N	2137.4	2137.1
Unimolecular (Dumbbell) Complexes: Linear						
3.4	^{HO}gcg-t_4-cgc-$A_{3',5'}$-cgc-t_4-gcgt$_{OH}$	88.0	11.1	N	6715.4	6711.1
3.5	^{HO}gcg-t_4-cgc-$A^{2',5'}$-cgc-t_4-gcgt$_{OH}$	98.8	6.1[b]	N	6715.4	6715.0
3.6	^{P}gcg-t_4-cgc-$A^{2',5'}$-cgc-t_4-gcgt$_{OH}$	66.7	2.9[b]	N L	6795.4 6777.4	6795.4 6775.9
3.7	^{HO}gcg-t_4-cgc-$A^{2',5'}$-cgc-t_4-gcgt$_{P}$	n.d.	9.3	N L	6795.4 6777.4	6796.7 6777.2
3.8	^{HO}tgcg-t_4-cgca-$A^{2',5'}$-cgc-t_4-gcgt$_{P}$	116.0	44.6	N L	7412.8 7394.8	7411.2 7392.2
3.9	^{HO}tgttgcg-t_4-cgcaaca-$A^{2',5'}$-cgc-t_4-gcgt$_{P}$	164.0	46.0	N L	9266.0 9247.9	9265.2 9250.8
3.10	^{HO}tgcg-t_4-cgcaaca-$A^{2',5'}$-cgc-t_4-gcgttgt$_{P}$	130.0	28.6	N L	9266.0 9247.9	9265.2 9250.8
Unimolecular (Dumbbell) Complexes: Branched						
3.11	^{HO}gcg-t_4-cgc-$A^{2,'5'}$(cgc-t_4-gcgt$_{P}$)$_{3',5'}$t$_{12}$	106	6.2	N	10445.7	n.d.
3.12	^{HO}tgcg-t_4-cgcaaca-$A^{2',5'}$(cgc-t_4-gcgttgt$_{P}$)$_{3',5'}$t$_{16}$	148.3	24.7	N L	14117.1 14099.1	14116.6 14097.9

Oligonucleotides were synthesized using a 1 μmol scale. Notation: small cap letters=deoxynucleotide residues, large cap letters=ribonucleotide residues. All oligonucleotides were purified by PAGE and desalted by SEC. [a]The crude and isolated yields reported are for the nicked compounds only. [b]Purification was conducted on only half of the crude material. P=phosphate. OH=hydroxyl. N=nicked compound. L=ligated compound. n.d.=not determined.

7.9.1. Synthesis of DNA dumbbell chimeras (Compounds 3.4-3.10)

Nicked DNA dumbbells were synthesized on a 1 μmol scale using the standard automated cycle with the noted exceptions (*vide infra*). Dumbbell sequences were selected so that comparisons could be made with a similar complex previously studied by Ashley and Kushlan, which merely lacked the 2'-rA insert.[269] All molecules were designed to favor exclusive formation of intramolecular duplexes.

The 5'-phosphorylated linear dumbbell (**3.6**) was synthesized by coupling 2-[2-(4,4'-dimethoxytrityloxy)ethylsulfonyl]ethyl-(β-cyanoethyl)-(N,N'-diisopropyl) phosphoramidite (ChemGenes) to the 5'-hydroxyl terminus of the grown DNA strand.[296] The phosphoramidite was dissolved at a concentration of 0.1 M, and a coupling time of 2 min was used. The 3'-phosphorylated linear dumbbells (Compounds **3.7-3.10**) were synthesized by coupling the phosphorylation phosphoramidite to the 5'-hydroxyl of any CPG-bound nucleoside and continuing synthesis of the strand in the normal 3'→5' direction. Upon deprotection under standard 3:1 NH_4OH/ethanol, a 3'-terminal phosphate was released.

All sequences were deprotected under standard conditions. The 2'-O-TBDMS (or 3'-O-TBDMS) protecting group was removed by treatment with TREAT-HF (100 μL) at room temperature for 48 h, followed by quenching with water and drying under vacuum. Nicked dumbbell oligonucleotides were analyzed by PAGE (20%, 8.3 M urea) and/or anion-exchange HPLC. Purification was by preparatory PAGE (20%, 8.3 M urea) followed by desalting on Sephadex G-25®. The sequences were characterized by negative mode MALDI-TOF-MS.

7.9.2. Synthesis of Branched DNA dumbbell (Compound 3.12)

The branched DNA dumbbell precursor was synthesized using the regiospecific method of Braich and Damha (section 7.3.4.2). The lariat precursor was designed such that the 5' and 2' extensions of the branched (Y-shaped) precursor are capable of

intramolecularly folding into a "dumbbell" complex, thereby aligning the reactive 5'-OH and 3'-phosphate for chemical ligation. Synthesis was conducted on a 1μmol scale using a low loading of dT-CPG (5-10 μmol/g). A 3'-phosphoryl group was introduced at the terminus of the 2'-branch extension using the commercial phosphorylation phosphoramidite (*vide supra*), however, an extended coupling time was used (10 min) owing to the increased sterics of extending the chain in the vicinity of the solid-support. Deprotection was conducted under standard conditions, and the bDNA analyzed by denaturing PAGE (12%, 8.3 M urea) and anion exchange HPLC. The desired band was resolved and purified from any failure sequences by PAGE (12%, 8.3 M urea), desalted on Sephadex G-25® and characterized by MALDI-MS: $[M]_{calc}$=14117.14, $[M-H]^-_{observed}$=14116.60. The overall isolated yield of the lariat-RNA precursor (**3.12**) was 24.7 A_{260} units (6.2% yield based on a 1 μmol synthesis).

7.9.3. Chemical Ligation with N-cyanoimidazole (ImCN)

Nicked dumbbell oligonucleotides (100 μM) containing either 5' or 3'-phosphorylated ends were dissolved in 100 mM 2-(N-morpholino)ethane sulfonic acid (MES, pH 6), 50 mM NaCl, and 100 mM $MnCl_2$ buffer. The solution was heated to 95°C, and allowed to intramolecularly associate into a dumbbell complex by cooling slowly at room temperature for 1 h, followed by cooling at 4°C overnight. An equal volume of N-cyanoimidazole (100 mM) dissolved in the same buffer was added to the samples and cooling continued for an additional 4-20 h. The oligonucleotides were precipitated directly from the reaction by adding 10 volumes of 2% $LiClO_4$ in reagent grade acetone, cooling on dry ice for *ca.* 30 min, followed by centrifugation at 14 Krpm for 10 min. The pellet was washed with cold acetone (2 × 0.25 mL), and dried. Ligated dumbbells were analyzed and purified by denaturing PAGE (20%, 8.3 M urea) and anion exchange HPLC. Prior to gel analysis, the samples were dissolved in gel loading buffer and heated at 95°C for 10 min in order to denature any intramolecular interactions that could result in the visualization of several species on the gel. Ligation of all compounds resulted in the formation of a single new product band, which migrated faster than the corresponding nicked precursor on a denaturing polyacrylamide gel. The yields of

ligated circles were determined by either densitometric analysis (UN-SCAN-IT Software: Silk Scientific) of the ligated dumbbell to unreacted precursor or by HPLC integration of the ensuing product peak. Products were characterized by MALDI-MS and by monitoring the increase in T_m between the nicked and ligated forms of the same sequence.

7.9.4. Chemical Ligation with Cyanogen Bromide (CNBr)

Phosphorylated (5' or 3') oligonucleotides (100 μM) were dissolved in 250 mM MES (pH 7.6) and 20 mM $MgCl_2$ buffer. Samples were denatured by heating to 95°C, and allowed to anneal into a dumbbell complex by slowly cooling at room temperature for 1 h, followed by cooling at 4°C overnight. The samples were further cooled on ice (0°C) for 15 minutes, at which time 5 M cyanogen bromide (CNBr) in acetonitrile (1/10 volume) was added. Ligation reactions utilizing CNBr were conducted in a well-ventilated fumehood. After 5 minutes, the oligonucleotides were precipitated directly from the reaction by adding 10 volumes of 2% $LiClO_4$ in acetone, cooling on dry ice for *ca.* 30 min, followed by centrifugation at 14 Krpm for 10 min. The pellet was washed with cold acetone (2 × 0.25 mL), and dried. Ligated dumbbells were analyzed and purified by denaturing PAGE (12-20%, 8.3 M urea) and anion exchange HPLC. Prior to gel analysis and purification, the samples were heat-treated as described above (CNIm ligation). Ligation of all compounds resulted in the formation of a single new product band, which migrated faster than the corresponding nicked precursor on a denaturing polyacrylamide gel. The yields of ligated circles and lariats were determined as described above (section 7.9.3) and characterized in the same manner.

7.10. CHAPTER 3: INTERMOLECULAR TEMPLATE-DIRECTED CHEMICAL LIGATION OF DNA AND RNA LARIATS

A complete list of the DNA and RNA sequences synthesized for this study is presented in **Table 7.3**. The crude and pure isolated yields as well as MALDI-TOF-MS characterization data are also reported.

Table 7.3: Crude yields, isolated yields and MALDI characterization of oligonucleotides used in the intermolecular template synthesis of lariat oligonucleotides.

Code	Sequence (5'→ 3')	Yield (ODU)			MW (g/mol)	
		Crude	Pure		Calc.	Found
DNA						
3.17	cgcaacaacaacgc	72.2	28.3		4210.83	4210.0
3.16	$tgttgcgt_{11}A^{2',5'}(t_{11}gcgttgt_P)_{3',5'}t_{20}$	104.6[a]	10.1[a]	**N**	17502.3	17513.6
				L	17484.3	17484.4
RNA						
3.18A	tGUUGCG	21.8	17.7		2195.4	2192.2
3.18B	$CGUUGt_P$	61.4	31.4		2275.4	2277.3
3.19	GCGUUGttGUUGCG	137.3	40.5		4452.7	4451.9
3.21	CGCAACAACAACGC	63.1	36.6		4434.8	4433.0
3.20	$tGUUGCGU_{11}A^{2',5'}(GU_{10}GCGUUGt_P)$ $_{3',5'}GU_{10}GCGUUGt_P$	89.4[a]	17.5[a]	**N**	17380.9	17382.3
				L	17362.9	17386.4

Oligonucleotides were synthesized using a 1 μmol scale. Notation: small cap letters=deoxynucleotide residues, large cap letters=ribonucleotide residues. P=terminal phosphate. Oligonucleotides were purified by PAGE and desalted by SEC. [a]The crude and isolated yields reported are for the nicked compounds only. N=nicked compound. L=ligated compound. n.d.=not determined.

7.10.1. Synthesis of the Branched DNA-Lariat Precursor (Compound 3.16)

The branched precursor for the intermolecular ligation of lariat DNA was synthesized using the regiospecific solid-phase methodology (section 7.3.4.2) on a 1 μmol scale (loading: 19 μmol/g). The sequence was designed to be devoid of any intramolecular complementarity, and the 5' and 2'-extensions were complementary to an intermolecular DNA or RNA template introduced during the chemical ligation step. Introduction of the

terminal 3'-phosphoryl group was conducted by reacting the 3'-hydroxyl of the branch extension with the chemical phosphorylation reagent; 2-[2-(4,4'-dimethoxytrityloxy)ethylsulfonyl]ethyl-(β-cyanoethyl)-(N,N'-diisopropyl) phosphoramidite (ChemGenes), however, an extended coupling time was used (10 min) owing to the increased sterics of extending the chain in the vicinity of the solid-support. The sequence was deprotected under standard conditions and purified by denaturing PAGE (12%, 7M urea) followed by desalting on Sephadex G-25® and characterization by MALDI-MS: $[M]_{calc}$=17502.31, $[M-2H+2Li]_{observed}$=17513.61. The overall isolated yield was 10.1 A_{260} units (2.1 % based on a 1 μmol synthesis).

7.10.2. Synthesis of the Branched RNA-Lariat Precursor (Compound 3.20)

The Y-shaped RNA-lariat precursor was synthesized using the convergent (non-regiospecific) methodology for bRNA synthesis on a 1 μmol scale (section 7.3.4.1). The 3'-phosphate was infixed by reacting a support-bound nucleoside (60 μmol/g, 1000 Å CPG) with the chemical phosphorylation reagent prior to oligonucleotide assembly. As with the DNA-lariat, the sequence was devised such that intramolecular complementarity was absent. Following deprotection under standard conditions, the 2'-TBDMS groups of the ribose sugars were released by treatment with TREAT-HF and precipitation from *n*-butanol as described (section 7.4). The branched RNA was purified by denaturing PAGE (12%, 7 M urea) and desalted on Sephadex G-25® and characterized by MALDI-MS:$[M]_{calc}$=17382.91, $[M-H]^-_{observed}$=17382.27. The isolated yield of branched lariat precursor was 17.5 A_{260} units (3.3 % based on a 1 μmol synthesis).

7.10.3. Template-Mediated Chemical Ligation of Lariats (Compounds 3.16 and 3.20)

DNA and RNA nicked branched precursors (**3.16** and **3.20**) and their corresponding linear splints (**3.17** and **3.21**) were dissolved in an equimolar ratio (100 μM each) in 0.25 M MES (pH 7.6), 20 mM $MgCl_2$ buffer. The samples were heated to 95°C for 10 minutes, cooled slowly to r.t., and left at 4°C overnight. Reactions were placed on ice

for 10 minutes at which time 1/10 volume of CNBr (5M in CH_3CN) was added, and the mixtures left on ice for an additional 5 minutes. The ligations were terminated by precipitation with 2% $LiClO_4$ in acetone (1 mL), placed at -20°C for 0.5-2 hours and the pellet recovered by centrifugation at 14 Krpm for 10 minutes. The supernatant was removed and the pellet dried. Ligated lariats were analyzed and purified by denaturing PAGE (12%, 8.3 M urea) and desalted on Sephadex G-25®. Prior to gel analysis and purification, the samples were heat-treated as described previously (CNIm ligation). The yields of ligated circles were determined by densitometric analysis (UN-SCAN-IT Software: Silk Scientific) of the ligated lariat to unreacted precursor. Compounds were characterized by MALDI-MS and the increase in T_m between the nicked and ligated entities as described previously.

7.11. CHAPTER 4: SELECTIVE AND POTENT INHIBITION OF HIV-1 REVERSE TRANSCRIPTASE (HIV-1 RT) BY AND RNA DUMBBELL

All experiments related to the inhibition of the RNase H activity of HIV reverse transcriptase were conducted with the collaboration of Dr. Kyung-Lyum Min of our research group. Nicked RNA dumbbells were synthesized in the identical fashion as the chimeric DNA dumbbells (section 7.9.1). Samples were purified and characterized as demonstrated in **Table 7.4**. Ligations were performed using CNBr as described previously (section 7.9.4) and the closed, circular compounds characterized by MALDI-TOF-MS and T_m analysis.

Table 7.4: Crude yields, isolated yields and MALDI characterization of RNA dumbbells.

Code	Sequence (5'→ 3')	Yield (ODU)			MW (g/mol)	
		Crude	Pure		Calc.	Found
4.1	^{HO}GGGAC(UUCG)GUCCAAAC(UUCG) GUUU$_P$	n.d.	22.8	N	7686.6	7707.5
4.2	^{HO}tGGGAC(UUCG)GUCCAAAAAC(UUCG) GUUUt$_P$	106.7[a]	37.6[a]	N	8953.4	8954.2
				L	8935.4	8931.0

Oligonucleotides were synthesized using a 1 μmol scale. Notation: small cap letters=deoxynucleotide residues, large cap letters=ribonucleotide residues. P=terminal phosphate. OH=hydroxyl. Oligonucleotides were purified by PAGE and desalted by SEC. [a]The crude and isolated yields reported are for the nicked compounds only. N=nicked compound. L=ligated compound. n.d.=not determined

7.11.1. Preparation of HIV-RT

The p66- and p51-kDA subunits of HIV-RT were prepared by Dr. Min of our laboratory by cloning into a pBAD/HisB prokaryotic expression vector (Invitrogen) between the XhoI and HindIII sites of the plasmid. The RT p66/p51 heterodimers and p51/p51 homodimers were purified as described previously by Fletcher and co-workers.[458]

7.11.2. Inhibition of the RNase H Activity of HIV-1 Reverse Transcriptase

The hybrid substrate for RNase H was prepared by labeling the 5'-hydroxy termini of the RNA sequence 5'-GAU CUG AGC CUG GGA GCU-3' by the transfer of ^{32}P from [γ-^{32}P]-ATP in a reaction catalyzed by bacteriophage T4 polynucleotide kinase as described previously (section 7.7.1). This labeled RNA was annealed to its complementary unlabeled DNA sequence, 5'-AGC TCC CAG GCT CAG ATC-3' to form the [^{32}P]-RNA/DNA hybrid substrate. Variable amounts of cold nicked and ligated RNA dumbbells were pre-incubated in 10 µL of 50 mM Tris-HCl (pH 8.0), containing 60 mM KCl, 2.5 mM $MgCl_2$ and 1.5 nM p51/p66 heterodimeric RT at 37°C for 15 minutes. The reactions were initiated by the addition of [^{32}P]-RNA/DNA hybrid duplex substrate (50 nM final concentration), and the individual assay tubes incubated a further 15 min. at 37°C. An equal volume of gel loading dye (98% deionized formamide containing 10 mM EDTA, 1 mg/mL bromophenol blue and 1 mg/mL xylene cyanol) was added to the samples and the reaction products denatured by heating at 100°C for 5 minutes. The degradation products were resolved on a 16% (19:1 crosslinking of acrylamide:bis-acrylamide) polyacrylamide sequencing gel (7 M urea) and visualized by autoradiography. The extent of cleavage of the 18-nt RNA portion of the RNA/DNA hybrid was determined quantitatively by densitometric analysis (UN-SCAN-IT software, Silk Scientific) of the disappearance of the full-length RNA and/or the appearance of any smaller degradation products. The IC_{50} values for RNA dumbbell inhibition of HIV-RT associated RNase H activity were calculated from plots of the residual undegraded 5'-[^{32}P]-RNA *versus* dumbbell concentration.

7.11.3. Inhibition of HIV-1 Reverse Transcriptase DNA Polymerase Activity

The assays described below were performed in order to determine if the inhibition of HIV-1 RT RNase H activity by RNA dumbbells was specific to this domain of the protein, without inhibiting its DNA-dependent or RNA-dependent DNA polymerase activity (*i.e.* HIV-1 RT mediated synthesis of DNA).

7.11.3.1. RNA-Dependent DNA Polymerase Activity Assay

The unlabeled, 30-nt RNA template, 5'-AUC UCU AGC AGA GGC GCC CGA ACA GGG ACA-3' (3-fold molar excess) was annealed to a 5'-[^{32}P]-end labeled complementary DNA primer; 5'-TGT CCC TGT TCG GGC GCC-3' in a separate reaction vessel. The RNA dumbbells (80 μM) were pre-incubated with the enzyme at room temperature for 20 minutes prior to reaction . Polymerase reactions were carried out in a 10 μL volume, containing 50 mM Tris-HCl (pH 8.0), 60 mM KCl, and 2.5 mM $MgCl_2$. The reaction was initiated by the addition of RNA template/5'-[^{32}P]-DNA primer complex and deoxynucleotide triphosphates (dNTPs, 200 nM final concentration of each) and incubated at 37°C for 15 minutes. The polymerase activity was deactivated by the addition of an equal volume of formamide loading dye (98% deionized formamide containing 10 mM EDTA, 1 mg/mL bromophenol blue and 1 mg/mL xylene cyanol) and denatured by heating at 100°C for 5 minutes prior to gel analysis (16%, 7M urea). The gel was visualized by autoradiography and the amount of DNA synthesized quantified by densitometric analysis using the UN-SCAN-IT software program.

7.11.3.2. DNA-Dependent DNA Polymerase Activity Assay

The ability for an RNA dumbbell to inhibit DNA synthesis from a DNA template stand was assessed using similar conditions to those used above in the RNA-dependent DNA polymerase assay. The 5'-[^{32}P]-DNA primer above was annealed to a 3-fold molar excess of DNA template, 5'-ATC TCT AGC AGA GGC GCC CGA ACA GGG ACA-3'. All other conditions for polymerization and analysis were identical to those described above.

7.11.4. *E. coli* and Human RNase H Inhibition Assays

In a subsequent experiment, the RNA dumbbell molecules were tested for their ability to inhibit either the *E. coli* or Human (type II) RNase H activities. RNase H mediated degradation assays were supplemented with 60 μM of cold RNA dumbbell under conditions identical to those used for HIV-RT RNase H. The degradation products were quantified from the autoradiogram using the UN-SCAN-IT software program.

7.11.5. UV-Crosslinking of HIV-1 RT RNase H Domain to an RNA Dumbbell

Homodimeric (p51/p51) and heterodimeric (p66/p51) HIV-1 RT enzymes (500 ng) were incubated with ligated RNA dumbbell 4.2 (50 pmol) in 50 mM Tris (pH 7.8), 50 mM KCl, and 5 mM $MgCl_2$ for 30 min at 37°C. The reaction mixtures were placed on ice and irradiated with a handheld UV-light (λ=254 nm) for 15 min. Samples were denatured by adding 2× sample loading buffer (4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 125 mM Tris, pH 6.8, and bromophenol blue) and heated at 100°C for 5 min. Protein complexes were partitioned on a 12% SDS-PAGE gel run at constant voltage (160 V). The gel was fixed with fixing solution (12% (w/v) trichloroacetic acid, 3.5 (w/v) 5-sulfosalicylic acid) for 30 min and then stained with Coomassie® Brilliant Blue G-perchloric acid solution (0.04% (w/v) Brilliant Blue G in 3.5% (w/v) perchloric acid); Sigma-Aldrich) for 60 min and finally rinsed with distilled water. Complexes were run alongside molecular weight markers consisting of ovalbumin (45 KDa), bovine serum albumin (66 KDa), phosphorylase B (97 KDa) and myosin (220 KDa).

7.12. CHAPTER 5: INHIBITION AND MODULATION OF PRE-mRNA SPLICING USING SYNTHETIC BRANCHED OLIGONUCLEOTIDES (bNAs)

Branched Oligonucleotides used in the ensuing studies were synthesized via the convergent solid-phase bNA synthetic methodology (section 7.3.4.1) on either 500Å CPG (<40 nts) or 1000Å CPG (>40 nts). Sequences were isolated and purified as described in **Tables 7.5** and **7.6** and characterized by MALDI-TOF-MS.

Table 7.5: Crude and isolated yields, purification techniques used and characterization of oligonucleotides used to study inhibition of *in vitro* pre-mRNA splicing in yeast and mammalian cell extracts

Code	Sequence (5'→ 3')	Yield (ODU)		Purification Method	MW (g/mol)	
		Crude	Pure		Calc.	Found
Linear Oligonucleotides						
5.10	tactaagtatg*c*	85.0	41.5	PAGE/SEC	3644.5	3643.5
5.11	UACUAAGUAUGU	53.1	7.5	HPLC/ppt.	3781.3	3781.1
5.12	UACUAAGUAUG*c*	48.9	22.3	HPLC/ppt.	3764.3	3763.5
5.13	**UACU**AAGUAUG*c*	53.1	22.4	HPLC/ppt.	3834.4	n.d.
V-Shaped Oligonucleotides						
5.14	$A^{2'5'}$(gtatg*c*)$_{3'5'}$gtatg*c*	54.4	14.8	PAGE/SEC	4005.7	4004.0
5.15	$A^{2'5'}$(GUAUG*c*)$_{3'5'}$GUAUG*c*	50.9	21.0	PAGE/SEC	4109.5	4113.5
Y-Shaped Oligonucleotides						
5.16	tacta$A^{2'5'}$(gtatg*c*)$_{3'5'}$gtatg*c*	61.7	12.9	PAGE/SEC	5529.7	5527.4
5.6	UACUAA$^{2'5'}$(GUAUGU)$_{3'5'}$GUAUGU	40.3	8.0	HPLC/ppt.	5719.5	5722.6
5.7	UACUAA$^{2'5'}$(GUAUG*c*)$_{3'5'}$GUAUG*c*	35.3	7.1	HPLC/ppt.	5685.5	5685.1
5.8	UACUAA$^{2'5'}$(GUAUG*ccc*)$_{3'5'}$GUAUG*ccc*	56.9	12.5	HPLC/ppt.	6842.2	6841.3
5.9	*c*UACUAA$^{2'5'}$(GUAUG*ccc*)$_{3'5'}$GUAUG*ccc*	43.9	4.5	HPLC/ppt.	7131.4	7133.9
5.17	*ccc*UACUAA$^{2'5'}$(GUAUG*ccc*)$_{3'5'}$GUAUG*ccc*	58.6	7.5	HPLC/ppt.	7709.8	7711.2
5.18	*ccc***UACUAA**$^{2'5'}$(**GUAUG***ccc*)$_{3'5'}$**GUAUG***ccc*	71.0	6.0	HPLC/ppt.	7815.0	n.d.
5.19	UACUA(aA)$^{2'5'}$(GUAUGU)$_{3'5'}$GUAUGU	53.6	4.9	PAGE/SEC	5719.5	5720.2
5.20	UACUA(aA)$^{2'5'}$(GUAUGU*c*)$_{3'5'}$GUAUGU*c*	36.7	2.3	PAGE/SEC	5685.5	n.d.

Oligonucleotides were synthesized using a 1 μmol scale. Notation: small cap letters=deoxynucleotide residues, large cap letters=ribonucleotide residues, large cap (bold)=2'-OMe-ribonucleotide residues, *c*=L-dC, aA=arabino adenosine. Samples purified by PAGE were partitioned on 20-24% acrylamide gels and those purified by HPLC were separated on an anion–exchange column as described. "SEC" refers to Sephadex desalting. "ppt." refers to precipitation after HPLC purification using *n*-propanol. n.d.=not determined.

7.12.1. Chapter 5.2.1: Inhibition of pre-mRNA Splicing in Yeast Nuclear Extract (*S. cerevisiae*)

Studies related to the inhibition of pre-mRNA splicing in *S. cerevisiae* were conducted with the cooperation of Dorian Anglin (University of Toronto) and in collaboration with Dr. James D. Friesen (University of Toronto, Banting and Best Dept. of Medical Research).

7.12.1.1. Preparation of Yeast pre-mRNA Splicing Extract

Yeast splicing extracts were prepared according to a modified method of Schultz and co-workers.[418] Splicing extracts were prepared from *S. cerevisiae* strain W3031A (*MAT*a, *ade2*-1, *his3*-11, *15 leu2*-3, *112 trp1*-1, *ura3*-1, *can1*-100).

Briefly, cells were inoculated into 2 × 1L YEPD (2% yeast extract/1% bactopeptone/2% dextrose) and cultured at 30°C for 12h with vigorous shaking. The cells were grown to an OD_{600} of 0.15 to 0.2, diluted to a volume of 1:10, and spun at 4 Krpm for 12 min in a Beckman J6-HC centrifuge. The weight of the cells was determined, and they were washed successively with cold water, and 1.3 volumes of extraction buffer [200mM Tris (pH 8.0), 0.39M $(NH_4)_2SO_4$, 10mM $MgSO_4$, 20% glycerol, 1mM EDTA (pH 7.9), 1mM DTT (added fresh)] supplemented with protease inhibitors [1mM phenylmethylsulfonyl fluoride (PMSF), 10mM benzamidine, 5 μg/mL leupeptin, 5 μg/mL pepstatin A]. The cell paste was scraped into a syringe, extruded into a tube filled with liquid nitrogen, and stored at –70°C.

Frozen cells (5 g) were ground into fine powder with a mortar and pestle under liquid nitrogen, and the powder thawed by placing over cold water. Extraction buffer (0.77-1 mL/1g of cells) was added and the extract spun in a 70.1 Ti rotor at 4°C, 35Krpm for 2 h to pellet the debris. All steps thereafter were performed at 4°C. The clear supernatant was retrieved, and solid $(NH_4)_2SO_4$ (0.337 g/mL) was added to the supernatant in small portions over the course of 0.5h, and mixed slowly by turning end over end for a further 0.5 h. The extract was pelleted at for 10-15 min at 25 Krpm, the pellet dissolved in

dialysis buffer (DB) [20mM HEPES (pH 7.5); 20% glycerol; 10mM $MgSO_4$; 10mM EGTA (pH 7.5); 5mM DTT (added fresh)] supplemented with protease inhibitor [1mM PMSF], and dialyzed against 2L DB for 8-12 h. The splicing extract was aliquoted and stored at –80°C until needed.

7.12.1.2. Yeast Nuclear Extract Oligonucleotide Stability Assays

Concentrated yeast pre-mRNA splicing extract was diluted 20-fold with 10 mM Tris-HCl, pH 7.8. Degradation assays were conducted in 12 µL total reaction volumes containing: 5'-end labeled branched oligonucleotide (*ca.* 5000 cpm/µL; 6 µL), diluted yeast extract (3 µL), 100 mM Tris-HCl, pH 7.8 (1.2 µL), 50 mM KH_2PO_4, pH 7.4 (1.2 µL) and the balance made up with sterile water. Reactions were incubated at 37°C, and a 4 µL aliquot removed at various time intervals (15, 30 and 60 min). The protein content was denatured and the nucleases inactivated by adding 4 µL of gel loading dye followed by heating the mixture at 70°C for 10 min. Samples were loaded alongside negative controls onto a denaturing polyacrylamide gel (16%, 7M urea), and run at 500 V for the first 30 min. to allow any protein-oligonucleotide complexes to completely enter the wells followed by 2000 V for the remaining time. The oligonucleotide degradation pattern was visualized by autoradiography (-20°C, 24h), and the oligonucleotide bands quantitated using the UN-SCAN-IT (Silk Scientific Corporation) image quantitating software.

7.12.1.3. Preparation of yeast actin pre-mRNA transcript

Full length yeast actin pre-mRNA for *in vitro* splicing assays was synthesized according to a modified procedure of Lin *et al.*[415] The T7 transcript was synthesized in 15 µL reactions containing the following: EcoRI linearized pGEM®9Zf(-) DNA vector containing the actin gene as template (1 µg/µL), 5 X T7 run off buffer (Amersham Pharmacia), 13 mM DTT, 1.1 mM each of ATP, GTP and CTP, 0.1 mM UTP, 10 mCi/mL [α-^{32}P]UTP, 40 units/µL RNAsin (Amersham Pharmacia Biotech), and 80 units/µL T7 RNA polymerase (Amersham). Transcription was conducted at room

temperature (25 °C) for 1.5 h, and the full-length transcripts purified using an RNeasy® Mini kit (Qiagen) according to manufacturer's protocol.

7.12.1.4. Yeast pre-mRNA Splicing Competitive Inhibition Assays

In vitro pre-mRNA splicing reactions were performed according to the method of Abelson and Ghetti[424] with slight modifications. Splicing reactions were effected in 10 μL total reaction volumes consisting of the following: 500 mM potassium phosphate buffer (1.2 μL), 30% PEG-8000 (1 μL), 150 mM $MgCl_2$ (0.2 μL), 20 mM ATP (0.5 μL), T7 pre-mRNA transcript (1000-10000 cpm/μL; 0.2 μL), yeast nuclear splicing extract (3 μL) and the remaining volume made up with sterilized water. For competitive inhibition assays, the splicing medium was complemented with cold exogenous branched RNA oligonucleotides at concentrations between 1-100 μM, prior to the addition of the yeast nuclear extract. Assays were conducted at room temperature for 20 min. in order to enable the visualization of any reaction products and intermediates (*i.e.* lariat-3'-exon, lariat and mRNA), and terminated through the addition of 10 μL of 1 × PK buffer [50 mM EDTA and 1% SDS] containing 2 mg/mL of proteinase K (Sigma). Subsequent to heating at 37°C for 15 min., the total RNA was isolated by phenol/chloroform extraction, equilibrated with 5 M sodium acetate (5 μL) and precipitated with 95% cold ethanol. Splicing reaction products were partitioned on a 5% denaturing polyacrylamide gel, and visualized and quantitated by phosphorimaging on a Molecular Dynamics® phosphorimager equipped with Image Quant® software.

7.12.2. Chapter 5.2.2: Inhibition of pre-mRNA Splicing in Mammalian Nuclear Extract (HeLa Cells)

HeLa nuclear extracts were generously donated by Dr. Andrew MacMillan (University of Alberta) and Dr. Benoit Chabot (University of Sherbrooke) and were used as received. All extracts were prepared according to the method of Dignam and contained

a final concentration of 100 mM KCl.[430] Extract was stored at -80°C and thawed on ice prior to use.

7.12.2.1. HeLa Nuclear Extract Oligonucleotide Stability Assays

The nuclease stability profiles of selected linear and branched oligonucleotides were conducted in the presence of HeLa nuclear extract under pre-mRNA splicing conditions (*i.e.* 25% extract). Substrate RNAs were radioactively labeled at their 5'-hydroxyl termini with a phosphate (^{32}P) prior to incubation (section 7.7.1). Degradation assays were accomplished in a 5 μL total volume and consisted of 15-50 × 10^3 CPM of 5'-end labeled linear or branched oligonucleotides, 25% (v/v) HeLa extract, 2 mM $MgCl_2$, 60 mM KCl (final concentration), RNase inhibitor (30 U), yeast phenylalanine tRNA (30 μg), 1 mM ATP and 5 mM creatine phosphate. The reactions were also supplemented with 10 mM KH_2PO_4 as a general phosphatase inhibitor in order to prevent loss of the 5'-radioactive label. The samples were incubated at 30°C for 15-90 minutes, diluted to 200 μL with sterile water and quenched with 200 μL PCI (phenol:chloroform:isosamyl alcohol; 25:24:1). The aqueous layer was removed and further extracted with chloroform (200 μL) and dried. Samples were partitioned on a 16% denaturing gel (7 M urea) and visualized by autoradiography. The amount of intact RNA was quantified using the UN-SCAN-IT software program (Silk Scientific) and the individual lanes normalized with respect to the reaction lacking any extract (t=0).

7.12.2.2. PCR Amplification of PIP85.B Splicing Substrate Gene

DNA plasmid containing the 234 nucleotide PIP85.B splicing gene was obtained from Dr. Andrew Macmillan (University of Alberta).[69,425] It contains a single adenosine in the branch region and a polypyrimidine tract (see Chapter 5; **Figure 5.13**). PCR amplification of the gene was conducted on a Minicycler™ programmable thermal controller (M.J. Research Inc, U.S.A.). The sequences of the PCR primers used were (1) M13F: 5'-CGC CAG GGT TTT CCC AGT CAC GAC-3' and (2) CQ27: 5'-AGC TTG CAT GCA GAG ACC-3'. The reaction was conducted in a total volume of 100 μL in a mini PCR tube containing: plasmid DNA (2.4 μg), primers M13F and CQ27 (2 μM of

each), 400 μM of dNTP's (equimolar mixture of dATP, dGTP, dCTP and dTTP), 10× *Taq* polymerase buffer (10 μL; Pharmacia), *Taq* polymerase enzyme (10 U) and the remaining volume made up with sterile water. Amplification was conducted for 30 cycles using an annealing temperature of 45°C for 30 seconds and a 90 second elongation time at 72°C. The dsDNA was purified from the reaction mixture by phenol/chloroform extraction. Briefly, the reaction volume was diluted to 250 mL with sterile water in a 1.5 mL microtube and an equal volume of PCI (25:24:1) added. The solution was mixed end over end for 5 min. and centrifuged at 10 Krpm for 10 min. at room temperature. The upper aqueous layer was removed and the phenol extraction repeated. An equal volume of chloroform was added, mixed well and centrifuged for 1 min. The aqueous layer was removed, diluted to 0.5 mL with sterile water and applied to a NAP-5® desalting column and eluted with sterile water (1 mL). The eluent was concentrated *in vacuo* and redissolved in 25 μL of water. The amplified DNA (8 μL) was mixed with 8 μL of gel loading dye (0.1% bromophenol blue, 15% glycerol), analyzed on a 2% agarose gel containing ethidium bromide (20 ng/mL) alongside a 1 Kb DNA ladder (1 μg) and visualized and photographed on a UV transilluminator.

7.12.2.3. Run-Off Transcription of PIP85.B Splicing Gene

The 234-nt RNA transcript was synthesized from the PCR amplified PIP85.B template using T7 RNA polymerase. Reactions were conducted in 25 μL volumes and consisted of: 5 μL of 5× T7 RNA polymerase buffer [200mM Tris-HCl (pH 7.9), 30mM $MgCl_2$, 50mM DTT, 50mM NaCl and 10mM spermidine], 100 mM DTT (2.5 μL), 10 mM solution of ATP, GTP and CTP (1.25 μL), 200 μM UTP (1.88 μL), RNase inhibitor (0.5 μL, 40 U/μL; Fermentas), [α-^{32}P]-UTP (2.5 μL, 50 μCi, 800 Ci/mmol; Amersham), PIP85.B DNA template (1 μL), T7 RNA Polymerase (1 μL, 20 U/μL; Fermentas) and sterile water to 25 μL. The transcription reaction was incubated at 37°C for 4 hours and cooled at -20°C overnight or on dry ice for 1 hour in order to inactivate the enzyme. The transcript was purified eitherby denaturing PAGE (8%, 7 M urea, 19:1 crosslinking) or using an RNeasy Mini Kit (Qiagen) according to manufacturer's specifications.

7.12.2.4. *In Vitro* pre-mRNA Splicing Reactions HeLa Extracts

Splicing of the PIP85.B pre-mRNA transcript was conducted under standard conditions with slight modifications in 5-10 μL reaction volumes.[93] Characteristically, reactions consisted of 25% (v/v) HeLa nuclear extract (containing 100 mM KCl), 2 mM $MgCl_2$, 60 mM KCl (final concentration), 1 mM ATP, 5 mM creatine phosphate, 0.4 U/mL RNase inhibitor, 50-100 × 10^3 CPM of internally radiolabelled transcript and the remaining volume was made up with sterile water. In the case of competitive inhibition assays, the reactions were supplemented with synthetic linear and branched oligonucleotides at concentrations ranging between 100 nM-20 μM. This was conducted by premixing the *hot*, radiolabelled pre-mRNA with *cold* oligonucleotide inhibitor. Samples were left on ice and the reactions initiated by the addition of the nuclear extract followed by heating to 30°C. Splicing reactions were incubated at 30°C for 30 minutes in order to resolve the pre-mRNA, mRNA, lariat-3'-exon and lariat intermediates. The protein content of the reactions was digested by the addition of 6 μL of a mixture containing Proteinase K (7.4 μg/ml final; Fermentas), PK stop buffer (2.6% SDS, 110 mM Tris, pH 8 and 110 mM EDTA), and yeast phenylalanine tRNA (1 μg/ml; Sigma) and heated at 55-60°C for 30 minutes. The reactions were quenched with 40 μL of sodium acetate (3M) and the volume brought up to 400 μL with sterile water. Individual reactions were extracted once with 200 μL PCI (25:24:1, phenol:chloroform:isoamyl alcohol) and once with 200 μL chloroform followed by precipitation with 2.5 volumes of ice-cold absolute ethanol. The samples were left on dry ice for a minimum of 2 hours, and centrifuged at maximum speed (14 Krpm) for 20 minutes. The supernatant ethanol was discarded and the pellet washed once with 50 μL of ice-cold ethanol. The pelleted RNA was dried in a Speed-Vac® concentrator for 2-5 minutes, and the amount of RNA quantitated on a Bio-Scan QC-2000 Benchtop Radioactivity Counter. Samples were dissolved in gel loading dye (2000 CPM/μL) and partitioned on a 15% (19:1 acrylamide:bis-acrylamide) containing 8 M urea. Gels were initially run at 500 V for 30 minutes in order to remove any residual salts from the samples, and then run at constant power (75 Watts) for 3.5-4 hours. Gels were autoradiographed as described previously (16-24 h, -20°C) and the extent of splicing

inhibition determined by quantitation of the lariat-3'-exon and lariat intermediate bands using the UN-SCAN-IT program.

7.12.3. Chapter 5.4: Modulation of Bcl-x Alternative Splicing By Linear and Branched Antisense Oligonucleotides

Experiments associated with the modulation of Bcl-x alternative splicing were performed in collaboration with Dr. Benoit Chabot and Jonathan Villemaire of the University of Sherbrooke. Large linear and branched synthetic oligonucleotides (antisense oligonucleotides; **Table 7.6**) were synthesized in our laboratory using solid-phase silyl phosphoramidite chemistry as described previously (section 7.3.4.1), purified by PAGE and desalted by size exclusion chromatography (SEC). HeLa nuclear extract for *in vitro* pre-mRNA splicing reactions was prepared using the method of Dignam and co-workers.[430]

The Bcl-x "mini-gene" employed in these assays (Bcl-x s*i*) is a viable construct consisting of key splicing regions in the full length Bcl-x gene. It comprises, (1) the 3'-end of exon 2 (371-nt) bearing the two alternative 5'-splice sites for Bcl-x_L and Bcl-x_S, (2) the 5'-end of exon 3 (138-nt) containing the conserved 3'-splice site, and (3) an intronic portion (between exons 2 and 3) consisting of the first 232-nt of the 5'-region and the last 182-nt of the 3'-region of the full length intron.

7.12.3.1. *In Vitro* Transcription of Bcl-x s*i*

The Bcl-x s*i* mini-gene was transcribed to its corresponding pre-mRNA by run-off transcription with bacteriophage T3 RNA polymerase. The EcoRI linearized plasmid containing the Bcl-x s*i* gene (1 μg) was incubated in the presence of transcription mixture, consisting of 1.25 μL of sterile water, 1.25 μL of 5× T3 polymerase buffer [1M Tris pH 7.5, 1 M $MgCl_2$, 1M spermidine and 1 M DTT], 0.625 μL of rNTPS (5 mM rATP, 5mM rCTP, and 1 mM rGTP), 0.625 μL of 100 mM UTP, 1.25 μL of [α-^{32}P]-UTP (800 Ci/mmol; Perkin Elmer Life Sciences), 0.25 μL of m^7G(5')ppp(5')G (Amersham Pharmacia Biotech), 0.25 μL RNase inhibitor (35 U/μL; Amersham

Table 7.6: Crude yields, isolated yields and MALDI characterization of oligonucleotides used in the modulation of Bcl-x study

Code	Sequence (5'→ 3')	Yield (ODU)		MW (g/mol)	
		Crude	Pure	Calc.	Found
Control RNAs					
5.23	GCC GCC GUU CUC CUG GAU CC	100.2	51.6	6270.8	6272.5
5.24	AUA GGC ACU GA GUU GGU AUG*a*	73.2	12.6	6777.1	6779.8
5.25	AUAGGCACUGA$^{2'5'}$(GUUGGUAUG*a*)$_{3'5'}$GUUGGUAUG*a*	61.3	7.2	10009.1	10007.2
Antisense with 5'-Overhangs					
5.26	AAU GUC UGC UAC UGG AAG AAU GCC GCC GUU CUC CUG GAU C *c*	79.7	38.2	13037.9	13037.3
5.27	UGG GUU UCU GAU AGG CAC UGA GCC GCC GUU CUC CUG GAU C *c*	87.1	21.6	13046.8	13039.1
5.28	UGG GUU UCU GAU AGG CAC UGA$^{2'5'}$(GCC GCC GUU CUC CUG GAU C *c*) $_{3'5'}$GCC GCC GUU CUC CUG GAU C *c*	83.2	8.5	19363.6	19361.4
Antisense with 3'-Overhangs					
5.29	GCC GCC GUU CUC CUG GAU CC AAU GUC UGC UAC UGG AAG AAU*c*	101.7	38.7	13343.1	13367.4
5.30	GCC GCC GUU CUC CUG GAU CC AUA GGC ACU GA GUU GGU AUG*a*	94.6	26.1	13093.9	13091.5
5.31	GCC GCC GUU CUC CUG GAU CC AAU GUC UGC UA$^{2'5'}$(AAU GUC UGC U*c*) $_{3'5'}$AAU GUC UGC U*c*	63.9	17.3	16730.0	16733.0
5.32	GCC GCC GUU CUC CUG GAU CC AAU GUC UGC UA$^{2'5'}$(GUU GGU AUG A*c*) $_{3'5'}$GUU GGU AUG A*c*	92.3	17.9	16890.1	16892.6
5.33	GCC GCC GUU CUC CUG GAU CC AUA GGC ACU GA$^{2'5'}$(AAU GUC UGC UAC UGG AAG AAU*c*)$_{3'5'}$AAU GUC UGC UAC UGG AAG AAU*c*	107.6	21.0	23990.5	23990.4
5.34	GCC GCC GUU CUC CUG GAU CC AUA GGC ACU GA$^{2'5'}$(GUU GGU AUG A*a*) $_{3'5'}$GUU GGU AUG A*a*	87.7	7.5	16341.8	16343.3

Oligonucleotides were synthesized using a 1 µmol scale, purified by PAGE and desalted by SEC. Underlined sequences are complementary to the 5'-splice site of Bcl-x_L. *a*=L-dA. *c*=L-dC

Pharmacia Biotech) and 0.25 μL of T3 RNA polymerase (40 U/μL; USB). The mixture was incubated at 30-37°C for 1 hour, and the radiolabeled transcript was purified on a 5% polyacrylamide gel in 1× TBE buffer. The RNA band of interest was excised from the gel, and extracted into 0.5 mL of extraction buffer (0.3 M sodium acetate and 0.2% SDS). The supernatant was extracted twice with PCI (25:24:1) and precipitated with two volumes of absolute ethanol. The pellet was air dried, resuspended in sterile water and stored at -80°C.

7.12.3.2. *In Vitro* Splicing Modulation of Bcl-x Expression

In vitro splicing reactions of the Bcl-x *s*i pre-mRNA were conducted in HeLa nuclear extracts as described previously.[417,450] Reactions were conducted in 15 μL total volumes and contained 5 μL HeLa extract, 1 μL pre-mRNA transcript (*ca.* 10000-15000 cpm), 0.5 μL of 12.5 mM ATP, 0.5 μL of 80 mM $MgCl_2$, 0.5 μL of 500 mM creatine phosphate, 2.5 μL of 13% PVA, 0.25 μL of 100 mM DTT, RNase inhibitor (1U/mL), creatine kinase (1U) and 4 μL of Buffer D [20 mM HEPES pH 7.9, 20% glycerol, 100 mM KCl, 0.2 mM EDTA, 0.5 mM DTT and 0.2 mM PMSF] and were supplemented with cold, antisense oligonucleotides (**5.23**-**5.34**) at concentrations ranging from 6.6-66 nM. The individual reactions were incubated at 30°C for 2 hours. The reactions were quenched with 450 □L of extraction solution (0.3 M NaOAc, 0.2% SDS and 20 μg tRNA carrier), PCI (25:24:1) extracted (500 □L) and precipitated with 2 volumes of absolute ethanol. The samples were stored at -80°C for 10 minutes, the spun down at 13 Krpm for 15 minutes. The RNA pellet was air-dried and resuspended in water.

7.12.3.3. RT-PCR of Bcl-x Splicing Products

The product RNAs produced from the splicing of the Bcl-x *s*i gene were amplified by RT-PCR. The RNA template must first be converted to a complementary DNA (cDNA) by way of the enzyme reverse transcriptase (RT). This cDNA mixture is then amplified by way of *Taq* DNA polymerase using multiple cycles of denaturation, annealing and elongation with specific DNA primers.

RT-PCR was accomplished using the commercially available Ready-To-Go™ RT-PCR beads (Amersham Pharmacia Biotech). The beads contained M-MuLV reverse transcriptase, RNase inhibitor, PCR buffer, dNTPs and *Taq* DNA polymerase in the appropriate amounts. The sequences of the DNA primers used were; Oligo T3'-5': 5'-GGG AAC AAA AGC TGG GTA CCG-3' and Bclx-2: 5'-TCA TTT CCG ACT GAA GAG TGA-3'. The beads were supplemented with Oligo T3'-5' (200 ng), Bclx-2 (400 ng), spliced RNAs (20 atmoles), [α-^{32}P]-dCTP (0.1 µL, 3000 Ci/mmol; Perkin Elmer Life Sciences) and water to a final volume of 50 µL. Amplification was conducted for 34 cycles using an annealing temperature of 55°C for 30 seconds and a 60 second elongation time at 72°C. The RNA components of the mixture were digested with RNase A (3 µL, 5 mg/mL) for 30 minutes at 37°C. The DNA assortment was then partitioned on a 4% (29:1 crosslinking) polyacrylamide gel (non-denaturing) run in 1× TBE buffer at 200 V. The gels were scanned on a Molecular Dynamics® Phosphorimager and the bands corresponding to the Bcl-x_L product (508 bp) and the Bcl-x_S product (319 bp) quantitated using ImageQuant® software.

REFERENCES

(1) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.

(2) Blackburn, G. M.; Gait, M. J. *Nucleic Acids in Chemistry and Biology*; Oxford University Press: New York, U.S.A., 1996.

(3) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737-738.

(4) Henry, C. M. In *Chemical and Engineering News*, March 10th, 2003, pp 49-60.

(5) Kelves, D. In *The Scientist*, January 13th, 2003; Vol. 17, pp 21-23.

(6) Avery, T. O.; Macleod, C.; McCarthy, M. *J. Exp. Med.* **1944**, *79*, 137-158.

(7) Gierer, A.; Schramm, G. *Nature* **1956**, *177*, 702-703.

(8) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 964-967.

(9) Chargaff, E. *Experientia* **1950**, *6*, 201-209.

(10) Moran, S.; Rex, X.-F.; Squire, R.; Kool, E. T. *J. Am. Chem. Soc.* **1997**, *119*, 2056-2057.

(11) Moran, S.; Rex, X.-F.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10506-10511.

(12) Morales, J. C.; Kool, E. T. *J. Am. Chem. Soc.* **1999**, *121*, 2323-2324.

(13) Delaney, J. C.; Henderson, P. T.; Lelquist, S. A.; Morales, J. C.; Essigmann, J. M.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4469-4473.

(14) Puglisi, J. D.; Tinoco, I. J. *Methods Enzymol.* **1989**, *180*, 304-325.

(15) Johnson, W. C. In *Circular Dichroism: Principles and Applications*; Nakanishi, K., Berova, N., Woody, R. W., Eds.; VCH Publishers: New York, 1994.

(16) Rodger, A.; Norden, B. *Circular Dichroism and Linear Dichroism*; Oxford University Press: New York, 1997.

(17) Westheimer, F. H. *Acc. Chem. Res.* **1968**, *1*, 70-78.

(18) Federoff, O. Y.; Salazar, M.; Reid, B. R. *J. Mol. Biol.* **1993**, *233*, 509-523.

(19) Ratmeyer, L.; Vinayak, R.; Zhong, Y. Y.; Zon, G. W.; Wilson, D. *Biochemistry* **1994**, *33*, 5298-5304.

(20) Salazar, M.; Champoux, J. J.; Reid, B. R. *Biochemistry* **1993**, *32*, 739-744.

(21) Salazar, M.; Federoff, O. Y.; Reid, B. R. *Biochemistry* **1996**, *35*, 8126-8135.

(22) Stryer, L. *Biochemistry*; 3rd Edition ed.; W. H. Freeman and Company: New York, 1988.

(23) Cech, T. R.; Zaug, A. J.; Grabowski, P. J. *Cell* **1981**, *27*, 487-496.

(24) Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. *Cell* **1982**, *31*, 147-157.

(25) Zaug, A. J.; Cech, T. R. *Science* **1986**, *231*, 470-475.

(26) Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. *Cell* **1983**, *35*, 849-857.

(27) Guerrier-Takada, C.; Altman, S. *Science* **1984**, *223*, 285-286.

(28) Altman, S.; Baer, M. F.; Bartkeiwicz, M.; Gold, H.; Guerrier-Takada, C.; Kirsebom, L. A.; Lumelsky, N.; Peck, K. *Gene* **1989**, *82*, 63-64.

(29) Cech, T. R. *J. Am. Med. Ass.* **1988**, *260*, 3030-3034.

(30) Noller, H. F.; Hoffarth, V.; Zimniak, L. *Science* **1992**, *256*, 1416-1419.

(31) Young, L. S.; Dunstan, H. M.; Witte, P. R.; Smith, T. P.; Ottonello, S.; Sprague, K. U. *Science* **1991**, *252*, 542-546.

(32) Temin, H. M.; Mizutani, S. *Nature* **1970**, *226*, 1211-1213.

(33) Baltimore, D. *Nature* **1970**, *226*, 1209-1211.

(34) Sharp, P. A. *Cell* **1985**, *42*, 397-400.

(35) *The RNA World: The Nature of RNA Suggests a Pre-Biotic RNA*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 1999.

(36) Chow, L. T.; Gelinas, R. E.; Broker, T. R.; Roberts, R. J. *Cell* **1977**, *12*, 1-8.

(37) Berget, S. M.; Moore, C.; Sharp, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 3171-3175.

(38) Sharp, P. A. *Cell* **1994**, *77*, 805-815.

(39) Faustino, N. A.; Cooper, T. A. *Genes Dev.* **2003**, *17*, 419-437.

(40) Cech, T. R. *Science* **1987**, *236*, 1532-1539.

(41) Michel, F.; Ferat, J. L. *Annu. Rev. Biochem.* **1995**, *64*, 435-461.

(42) Ferat, J. L.; Michel, F. *Nature* **1993**, *364*, 289-290.

(43) Jacquier, A. *Trends Biochem. Sci.* **1990**, *15*, 351-354.

(44) Weiner, A. M. *Cell* **1993**, *72*, 161-164.

(45) Nilsen, T. W. *Exp. Parasitol.* **1989**, *69*, 413-416.

(46) Eul, J.; Graessmann, M.; Graessmann, A. *EMBO J.* **1995**, *14*, 3226-3235.

(47) Anderson, K.; Moore, M. J. *Science* **1997**, *276*, 1712-1716.

(48) Garcia-Blanco, M. A.; Puttaraju, M.; Mansfield, S. G.; Mitchell, L. G. *Gene Ther. Reg.* **2000**, *1*, 141-163.

(49) Madhani, H. D.; Guthrie, C. *Annu. Rev. Genet.* **1994**, *28*, 1-26.

(50) Brody, E.; Abelson, J. *Science* **1985**, *228*, 965-967.

(51) Grabowski, P. J.; Seiler, S. R.; Sharp, P. A. *Cell* **1985**, *42*, 345-353.

(52) Zhou, Z.; Licklider, L. J.; Gygi, S. P.; Reed, R. *Nature* **2002**, *419*, 182-185.

(53) Zhang, M.; Marr, T. *Nucl. Acids. Res.* **1994**, *22*, 1750-1759.

(54) Moore, M. J.; Query, C. C.; Sharp, P. A. In *The RNA World*; Gesteland, R., Atkins, J., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 1993.

(55) Staley, J. P.; Guthrie, C. *Cell* **1998**, *92*, 315-326.

(56) Reed, R.; Maniatis, T. *Genes Dev.* **1988**, *2*, 1268-1276.

(57) Zhuang, Y.; Goldstein, A. M.; Weiner, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 2752-2756.

(58) Hayles, J.; Nurse, P. *Annu. Rev. Genet.* **1992**, *26*, 373-402.

(59) Gaur, R. K.; Valcarcel, J.; Green, M. R. *RNA* **1995**, *1*, 407-417.

(60) Abovich, N.; Liao, X.; Rosbash, M. *Genes Dev.* **1994**, *8*, 843-854.

(61) Lerner, M. R.; Boyle, J. A.; Mount, S. M.; Wolin, S. L.; Steitz, J. A. *Nature* **1980**, *282*, 220-224.

(62) Mount, S. M.; Pettersson, I.; Hinterberger, M.; Karmas, A.; Steitz, J. A. *Cell* **1983**, *33*, 509-518.

(63) Kramer, A.; Keller, W.; Appel, B.; Luhrmann, R. *Cell* **1984**, *38*, 299-307.

(64) Yean, S.-L.; Lin, R.-J. *Gene Expr.* **1996**, *5*, 301-313.

(65) Black, D. L.; Chabot, B.; Steitz, J. A. *Cell* **1985**, *42*, 737-750.

(66) Parker, R.; Siciliano, P. G.; Gurthrie, C. *Cell* **1987**, *49*, 229-239.

(67) Wu, J.; Manley, J. *Genes Dev.* **1989**, *3*, 1553-1561.

(68) Zhuang, Y.; Weiner, A. *Genes Dev.* **1989**, *3*, 1545-1552.

(69) Query, C. C.; Moore, M. J.; Sharp, P. A. *Genes Dev.* **1994**, *8*, 587-597.

(70) Zavanelli, M. I.; Britton, J. S.; Igel, A. H.; Ares, M. J. *Mol. Cell. Biol.* **1994**, *14*, 1689-1697.

(71) Nilsen, T. W. In *RNA Strucuture and Function*; Simons, R., Grunberg-Manago, M., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 1998.

(72) Chabot, B.; Black, D. L.; LeMaster, D. M.; Steitz, J. A. *Science* **1985**, *230*, 1344-1349.

(73) Newman, A. J. *EMBO J.* **1997**, *16*, 5797-5800.

(74) Ast, G.; Weiner, A. M. *RNA* **1997**, *3*, 371-381.

(75) Guthrie, C. *Science* **1991**, *253*, 157-163.

(76) Datta, B.; Weiner, A. M. *Nature* **1991**, *352*, 821-824.

(77) Madhani, H. D.; Guthrie, C. *Cell* **1992**, *71*, 803-817.

(78) Wassarman, D. A.; Steitz, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 7139-7143.

(79) Wu, J.; Manley, J. L. *Nature* **1991**, *352*, 818-821.

(80) Collins, C. A.; Guthrie, C. *Nature Struc. Biol.* **2000**, *7*, 850-854.

(81) Villa, T.; Pleiss, J. A.; Guthrie, C. *Cell* **2002**, *109*, 149-152.

(82) Valadkhan, S.; Manley, J. L. *RNA* **2000**, *6*, 206-219.

(83) Valadkhan, S.; Manley, J. L. *Nature* **2001**, *413*, 701-707.

(84) Valadkhan, S.; Manley, J. L. *RNA* **2003**, *9*, 892-904.

(85) Konarska, M. M.; Sharp, P. A. *Cell* **1986**, *46*, 845-855.

(86) Konarska, M. M.; Sharp, P. A. *Cell* **1987**, *49*, 763-774.

(87) Stevens, S. W.; Abelson, J. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 7226-7231.

(88) Stevens, S. W.; Ryan, D. E.; Ge, H. Y.; Moore, R. E.; Young, M. K.; Lee, T. D.; Abelson, J. *Mol. Cell* **2002**, *9*, 31-44.

(89) Jurica, M. S.; Moore, M. J. *Mol. Cell* **2003**, *12*, 5-14.

(90) Nilsen, T. W. *Mol. Cell* **2002**, *9*, 8-9.

(91) Konarska, M. M.; Sharp, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 5459-5462.

(92) Padgett, R. A.; Hardy, S. F.; Sharp, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 5230-5234.

(93) Grabowski, P. J.; Padgett, R. A.; Sharp, P. A. *Cell* **1984**, *37*, 415-427.

(94) Ruskin, B.; Krainer, A. R.; Maniatis, T.; Green, M. R. *Cell* **1984**, *38*, 317-331.

(95) Maschhoff, K.; Padgett, R. *Nucl. Acids. Res.* **1993**, *21*, 5456-5462.

(96) Moore, M. J.; Sharp, P. A. *Nature* **1993**, *365*, 364-368.

(97) McSwiggen, J. A.; Cech, T. R. *Science* **1989**, *244*, 679-683.

(98) Rajagopal, J.; Doudna, J. A.; Szostak, J. *Science* **1989**, *244*, 692-694.

(99) Wallace, J. C.; Edmonds, M. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 950-954.

(100) Domdey, H.; Apostol, B.; Lin, R.-J.; Newman, A.; Brody, E.; Abelson, J. *Cell* **1984**, *39*, 611-621.

(101) Padgett, R. A.; Konarska, M. M.; Grabowski, P. J.; Hardy, S. F.; Sharp, P. A. *Science* **1984**, *225*, 898-903.

(102) Konarska, M. M.; Grabowski, P. J.; Padgett, R. A.; Sharp, P. A. *Nature* **1985**, *313*, 552-557.

(103) Freyer, G. A.; Arenas, J.; Perkins, K. K.; M., F. H.; Pick, L.; Young, B.; Roberts, R. J.; Hurwitz, J. *J. Biol. Chem.* **1987**, *262*, 4267-4273.

(104) Hornig, H.; Aebi, M.; Weissmann, C. *Nature* **1986**, *324*, 589-591.

(105) Ruskin, B.; Greene, J. M.; Green, M. R. *Cell* **1985**, *41*, 833-844.

(106) Langford, C. J.; Klinz, F.-J.; Donath, C.; Gallwitz, D. *Cell* **1984**, *36*, 645-653.

(107) Remaud, G.; Balgobin, N.; Glemarec, C.; Chattopadhyaya, J. *Tetrahedron* **1989**, *45*, 1537-1548.

(108) Query, C. C.; Strobel, S. A.; Sharp, P. A. *Eur. Mol. Biol. J.* **1996**, *15*, 1392-1402.

(109) Ruskin, B.; Green, M. R. *Science* **1985**, *229*, 135-140.

(110) Ruskin, B.; Green, M. R. *Cell* **1985**, *43*, 131-142.

(111) Chabot, B.; Steitz, J. *Mol. Cell. Biol.* **1987**, *7*, 281-293.

(112) Arenas, J.; Hurwitz, J. *J. Biol. Chem.* **1987**, *262*, 4274-4279.

(113) Chapman, K. B.; Boeke, J. D. *Cell* **1991**, *65*, 483-492.

(114) Furuichi, T.; Dhundale, A.; Inouye, M.; Inouye, S. *Cell* **1987**, *48*, 47-53.

(115) Nam, K.; Hudson, R. H. E.; Chapman, K. B.; Ganeshan, K.; Damha, M. J.; Boeke, J. D. *J. Biol. Chem.* **1994**, *269*, 20613-20621.

(116) Ooi, S. L.; Dann, C. I.; Nam, K.; Leahy, D. J.; Damha, M. J.; Boeke, J. D. *Methods Enzymol.* **2001**, *342*, 233-250.

(117) Liscio, A. , M.Sc. Thesis, McGill University, 2000.

(118) Nam, K.; Lee, G.; Trambley, J.; Devine, S. E.; Boeke, J. D. *Mol. Cell. Biol.* **1997**, *17*, 809-818.

(119) Petfalski, E.; Dandekar, T.; Henry, Y.; Tollervey, D. *Mol. Cell. Biol.* **1998**, *18*, 1181-1189.

(120) Ooi, S. L.; Samarsky, D. A.; Fournier, M. J.; Boeke, J. D. *RNA* **1998**, *4*, 1096-1110.

(121) Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.*et al. Nature* **2001**, *409*, 860-921.

(122) Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.*et al. Science* **2001**, *291*, 1304-1351.

(123) Modrek, B.; Lee, C. *Nature Gene.* **2002**, *30*, 13-19.

(124) Chabot, B. *Trends Gen.* **1996**, *12*, 472-478.

(125) Smith, C. W. J.; Valcarcel, J. *T.I.B.S.* **2000**, *25*, 381-388.

(126) Grabowski, P. J. *Cell* **1998**, *92*, 709-712.

(127) Cline, T. W.; Meyer, B. J. *Annu. Rev. Genet* **1996**, *30*.

(128) Jiang, Z.; Wu, J. *Proc. Soc. Exp. Biol. Med.* **1999**, *220*, 64-72.

(129) Lopez, A. J. *Annu. Rev. Genet.* **1998**, *32*, 279-305.

(130) Herbert, A.; Rich, A. *Nature Genet.* **1999**, *21*, 265-269.

(131) Zahler, A.; Lane, W.; Stolk, J.; Roth, M. *Genes Dev.* **1992**, *6*, 837-847.

(132) Fu, X. D. *RNA* **1995**, *1*, 663-680.

(133) Misteli, T. *Curr. Biol.* **1999**, *9*, 198-200.

(134) Berget, S. M. *J. Biol. Chem.* **1995**, *270*, 2411-2414.

(135) Dominski, Z.; Kole, R. *Mol. Cell. Biol.* **1992**, *12*, 2108-2114.

(136) Boise, L. H.; Gonzalez-Garcia, M.; Postema, C. E.; Ding, L.; Lindsten, T.; Turka, L. A.; Mao, X.; Nunez, G.; Thompson, C. B. *Cell* **1993**, *74*, 597-608.

(137) Zhou, M.; Demo, S. D.; McClure, T. N.; Crea, R.; Bitler, C. M. *J. Biol. Chem.* **1998**, *273*, 11930-11936.

(138) Tao, W.; Kurschnew, C.; Morgan, J. I. *J. Biol. Chem.* **1998**, *273*, 23704-23708.

(139) Bennett, C. F.; Dean, N. M.; Monia, B. P.; Nickoloff, B. J.; Zhang, Q. Q., U.S.A., **2001**.

(140) Mercatante, D. R.; Bortner, C. D.; Cidlowski, J. A.; Kole, R. *J. Biol. Chem.* **2001**, *276*, 16411-16417.

(141) Taylor, J. K.; Zhang, Q. Q.; Wyatt, J. R.; Dean, N. M. *Nature Biotech.* **1999**, *17*, 1097-1100.

(142) Mercatante, D.; Kole, R. *Pharmacol. Ther.* **2000**, *85*, 237-243.

(143) Sazani, P.; Kole, R. *Prog. Mol. Subcell. Biol.* **2003**, *31*, 217-239.

(144) Khorana, H. G.; Todd, A. R. *J. Chem. Soc. Abs.* **1953**, 2257-2260.

(145) Gilham, P. T.; Khorana, H. G. *J. Am. Chem. Soc.* **1958**, *80*, 6212-6222.

(146) Nishimura, S.; Jacob, H.; Khorana, H. G. *Proc. Natl. Acad. Sci. U.S.A.* **1964**, *52*, 1494-1501.

(147) Agrawal, K. L.; Buchi, H.; Caruthers, M. H.; Gupta, N.; Khorana, H. G.*et al. Nature* **1970**, *227*, 27-34.

(148) Letsinger, R. L.; Mahadevan, V. *J. Am. Chem. Soc.* **1966**, *88*, 5319-5324.

(149) Letsinger, R. L.; Mahdevan, V. *J. Am. Chem. Soc.* **1965**, *87*, 3526-3527.

(150) Letsinger, R. L.; Ogilvie, K. K. *J. Am. Chem. Soc.* **1967**, *89*, 4801-4803.

(151) Letsinger, R. L.; Ogilvie, K. K. *J. Am. Chem. Soc.* **1969**, *91*, 3350-3355.

(152) Tener, G. M. *J. Am. Chem. Soc.* **1961**, *83*, 159-168.

(153) Reese, C. B. *Tetrahedron* **1978**, *34*, 3143-3179.

(154) Letsinger, R. L.; Finnan, J. L.; Heavner, G. A.; Lunsford, W. B. *J. Am. Chem. Soc.* **1975**, *97*.

(155) Letsinger, R. L.; Lunsford, W. B. *J. Am. Chem. Soc.* **1976**, *98*, 3655-3661.

(156) Matteucci, M. D.; Caruthers, M. H. *Tetrahedron Lett.* **1980**, *21*, 719-722.

(157) Daub, G. W.; van Tamelen, E. E. *J. Am. Chem. Soc.* **1977**, *99*, 3526-3528.

(158) Ogilvie, Kelvin K; Nemer, M. J. *Tetrahedron Lett.* **1980**, *21*, 4145-4148.

(159) Ogilvie, Kelvin K.; Nemer, M. J. *Tetrahedron Lett.* **1980**, *21*, 4159-4162.

(160) Beaucage, S. L.; Caruthers, M. H. *Tetrahedron Lett.* **1981**, *22*, 1859-1862.

(161) McBride, L. J.; Caruthers, M. H. *Tetrahedron Lett.* **1983**, *24*, 245-248.

(162) Adlvaredo-Urbina, G.; Sathe, G. M.; Liu, W. C.; Gillen, M. F.; Duck, P. D.; Bender, R.; Ogilvie, K. K. *Science* **1981**, *214*, 270-274.

(163) Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149-2154.

(164) Letsinger, R. L.; Kornet, M. J.; Mahadevan, V.; Jerina, D. M. *J. Am. Chem. Soc.* **1964**, *86*, 5163-5165.

(165) Pon, R. T.; Usman, N.; Ogilvie, K. K. *Biotechniques* **1988**, *6*, 768-775.

(166) Damha, M. J.; Giannaris, P. A.; Zabarylo, S. V. *Nucl. Acids Res.* **1990**, *18*, 3813-3821.

(167) Pon, R. T.; Shuyuan, Y. *Nucl. Acids Res.* **1997**, *25*, 3629-3635.

(168) Pon, R. T.; Yu, S. *Tetrahedron Lett.* **1997**, *38*, 3327-3330.

(169) Schaller, H.; Weimann, G.; Lerch, B.; Khorana, H. G. *J. Am. Chem. Soc.* **1963**, *85*, 3821-3827.

(170) Hakimelahi, G. H.; Proba, Z. A.; Ogilvie, K. K. *Can. J. Chem.* **1982**, *60*, 1106-1113.

(171) Ti, G. S.; Gaffney, B. L.; Jones, R. A. *J. Am. Chem. Soc.* **1982**, *104*, 1316-1319.

(172) Ogilvie, Kelvin K.; Sadana, K. L.; Thompson, E. A.; Quilliam, M. A.; Westmore, J. B. *Tetrahedron Lett.* **1974**, *15*, 2861-2863.

(173) Ogilvie, K. K.; Beaucage, S. L.; Schiffman, A. L.; Theriault, N. Y.; Sadana, K. L. *Can. J. Chem.* **1978**, *56*, 2768-2680.

(174) Corey, E. J.; Venkateswarlu, A. *J. Am. Chem. Soc.* **1972**, *94*, 6190-6191.

(175) Scaringe, S. A.; WIncott, F. E.; Caruthers, M. H. *J. Am. Chem. Soc.* **1998**, *120*, 11820-11821.

(176) Dahl, B. H.; Nielsen, J.; Dahl, O. *Nucl. Acids Res.* **1987**, *15*, 1729-1743.

(177) Berner, S.; Muhlegger, K.; Seliger , H. *Nucl. Acids Res.* **1989**, *17*, 853-864.

(178) Temsamani, J.; Kubert, M.; Agrawal, S. *Nucl. Acids Res.* **1995**, *23*, 1841-1844.

(179) Farrance, I. K.; Eadie, J. S.; Ivarie, R. *Nucl. Acids Res.* **1989**, *17*, 1231-1245.

(180) Eadie, J. S.; Davidson, D. S. *Nucl. Acids Res.* **1987**, *15*, 8333-8349.

(181) Pon, R. T.; Usman, N.; Damha, M. J.; Ogilvie, K. K. *Nucl. Acids Res.* **1986**, *14*, 6453-6470.

(182) Gasparutto, D.; Livache, T.; Bazin, H.; Duplaa, A.-M.; uy, A.; Khorlin, A.; Molko, D.; Roget, A.; Teoule, R. *Nucl. Acids Res.* **1992**, *20*, 5159-5166.

(183) Jarvinen, P.; Oivanen, M.; Lonnberg, H. *J. Org. Chem.* **1991**, *56*.

(184) Nilsen, T. W. *Infectious Agents and Disease* **1992**, *1*, 212-218.

(185) Yee, T.; Furuichi, T.; Inouye, S.; Inouye, M. *Cell* **1984**, *38*.

(186) Damha, M. J.; Pon, R. T.; Ogilvie, K. K. *Tetrahedron Lett.* **1985**, *26*, 4839-4842.

(187) Damha, M. J.; Ogilvie, K. K. *J. Org. Chem.* **1988**, *53*, 3710-3722.

(188) Sekine, M.; Hata, T. *J. Am. Chem. Soc.* **1985**, *107*, 5813-5815.

(189) Kierzek, R.; Kopp, D. W.; Edmonds, M.; Caruthers, M. H. *Nucl. Acids Res.* **1986**, *14*, 4751-4764.

(190) Hayakawa, Y.; Noboria, T.; Noyori, R.; Imai, J. *Tetrahedron Lett.* **1987**, *28*, 2623-2626.

(191) Fourrey, J. L.; Varenne, J.; Fontaine, C.; Guittet, E.; Yang, Z. W. *Tetrahedron Lett.* **1987**, *28*, 1769-1771.

(192) Huss, S.; Gosselin, G.; Imbach, Jean-Louis *Tetrahedron Lett.* **1987**, *28*, 415-418.

(193) Zhou, X.-. ; Remaud, G.; Chattopadhyaya, J. *Tetrahedron* **1988**, *44*, 6471-6489.

(194) Foldesi, A.; Balgobin, N.; Chattopadhyaya, J. *Tetrahedron Lett.* **1989**, *30*, 881-884.

(195) Pathak, T.; Chattopadhyaya, J. *Acta. Chem. Scan.* **1985**, *B39*, 799-806.

(196) Huss, S.; Gosselin, G.; Stawinski, J.; Stromberg, R.; Imbach, J.-L. *Nuclesosides and Nucleotides* **1988**, *7*, 321-337.

(197) de Rooij, J. F.; Wille-Hazeleger, G.; Burgers, P. M.; van Boom, J. H. *Nucl. Acids Res.* **1979**, *6*, 237.

(198) Reese, C. B.; Skone, P. A. *Nucl. Acids Res.* **1985**, *13*, 5215-5231.

(199) Huss, S.; Gosselin, G.; Imbach, J.-L. *J. Org. Chem.* **1988**, *53*, 499-506.

(200) Sekine, Mitsuo; Heikkila, J.; Hata, T. *Tetrahedron Lett.* **1987**, *28*, 5691-5694.

(201) Balgobin, N.; Foldesi, A.; Remaud, G.; Chattopadhyaya, J. *44* **1988**, *6929-6939*.

(202) Hall, R. H.; Todd, A. R.; Webb, R. F. *J. Am. Chem. Soc.* **1957**, *3291*, 3291-3296.

(203) Damha, M. J.; Ogilvie, K. K. In *Protocols for Oligonucleotides and Analogs*; Agrawal, S., Ed.; Humana Press Inc.: Totowa, N.J., 1993; Vol. 20.

(204) Damha, M. J.; Zabarylo, S. *Tetrahedron Lett.* **1989**, *30*, 6295-6298.

(205) Zabarylo, S. , M. Sc. Thesis, University of Toronto, 1991.

(206) Damha, M. J.; Ganeshan, K.; Hudson, R. H. E.; Zabarylo, S. V. *Nucl. Acids Res.* **1992**, *20*, 6565-6573.

(207) Hudson, R. H. E.; Damha, M. J. *J. Am. Chem. Soc.* **1993**, *115*, 2119-2124.

(208) Hudson, R. H. E. , Ph.D. Thesis, University of Toronto, 1995.

(209) Hudson, R. H. E.; Robidoux, S.; Damha, M. J. *Tetrahedron Lett.* **1998**, *39*, 1299-1302.

(210) Hudson, R. H. E.; Uddin, A. H.; Damha, M. J. *J. Am. Chem. Soc.* **1995**, *117*, 12470-12477.

(211) Uddin, A. H.; Piunno, P. A. E.; Hudson, R. H. E.; Damha, M. J.; Krull, U. J. *Nucl. Acids Res.* **1997**, *25*, 4139-4146.

(212) Braich, R. S. , Ph.D. Thesis, McGill University, 1999.

(213) Carriero, S.; Mangos, M. M.; Agha, K. A.; Noronha, A. M.; Damha, M. J. *Nucleosides, Nucleotides and Nucleic Acids* **2003**, *in press*.

(214) Ganeshan, K.; Tadey, T.; Nam, K.; Braich, R.; Purdy, W. C.; Boeke, J. D.; Damha, M. J. *Nucleosides & Nucleotides* **1995**, *14*, 1009-10013.

(215) Ganeshan, K. , Ph.D. Thesis, McGill University, 1995.

(216) Sproat, B. S.; Beijer, B.; Grotli, M.; Ryder, U.; Morand, K. L.; Lamond, A. I. *J. Am. Chem. Soc. Perkin Trans.* **1994**, *1*, 419-431.

(217) Grotli, M.; Sproat, B. S. *J. Chem. Soc., Chem. Comm.* **1995**, *4*, 495-497.

(218) Grotli, M.; Eritja, R.; Sproat, B. *Tetrahedron* **1997**, *53*, 11317-11346.

(219) von Buren, M.; Petersen, G. V.; Rasmussen, K.; Brandenburg, G.; Wengel, J. *Tetrahedron* **1995**, *51*, 8491-8506.

(220) Brandenburg, G.; Petersen, G. V.; Rasmussen, K.; Wengel, J. *Bioorg. Med. Chem. Lett.* **1995**, *5*, 791-794.

(221) Braich, R. S.; Damha, M. J. *Bioconj. Chem.* **1997**, *8*, 370-377.

(222) Damha, M. J.; Braich, R. S. *Tetrahedron Lett.* **1998**, *39*, 3907-3910.

(223) Sund, C.; Agback, P.; Chattopadhyaya, J. *Tetrahedron* **1991**, *47*, 9659-9674.

(224) Sund, C.; Agback, P.; Chattopadhyaya, J. *Tetrahedron* **1993**, *49*, 649-668.

(225) Agback, P.; Glemarec, C.; Yin, L.; Sandstrom, A.; Plavec, J.*et al. Tetrahedron Lett.* **1993**, *34*, 3929-3932.

(226) Reese, C. B.; Song, Q. *Nucl. Acids Res.* **1999**, *27*, 2672-2681.

(227) Hannoush, R. N. , Ph. D. Thesis, McGill University, 2002.

(228) Carriero, S.; Braich, R. S.; Hudson, R. H. E.; Anglin, D.; Friesen, J. D.; Damha, M. J. *Nucleosides, Nucleotides and Nucleic Acids* **2001**, *20*, 873-877.

(229) Reilly, J. D.; Freeman, S. K.; Melhem, R. F.; Kierzek, R.; Caruthers, M. H.; Edmonds, M.; Munns, T. W. *Anal. Biochem.* **1990**, *185*, 125-130.

(230) Hudson, R. H. E.; Ganeshan, K.; Damha, M. J. In *Carbohydrate Modifications in Antisense Research*; Cook, P. D., Sanghvi, Y. S., Eds.; American Chemical Society: Washington, DC., 1994; Vol. 580.

(231) Sorensen, M. D.; Meldgaard, M.; Rajwanshi, V. K.; Wengel, J. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1853-1856.

(232) Barbato, S.; De Napoli, L.; Mayol, L.; Piccialli, G.; Santacroce, G. *Tetrahedron Lett.* **1987**, *28*, 5727-5728.

(233) Carriero, S.; Damha, M. J. In *Current Protocols in Nucleic Acid Chemistry*; Beaucage, S., Ed.; John Wiley & Sons:, 2002; Vol. 4.14.

(234) Tan, W.; Iyer, R. P.; Jiang, Z.; Agrawal, S., U.S. Patent #5,631,361, **1997**.

(235) Usman, N.; Ogilvie, K. K.; Jiang, M. Y.; Cedergren, R. J. *J. Am. Chem. Soc.* **1987**, *109*, 7845-7854.

(236) Pon, R. T.; Yu, S.; Sanghvi, Y. S. *Bioconj. Chem.* **1999**, *10*, 1051-1057.

(237) Padgett, R. A.; Grabowski, P. J.; Konarska, M. M. M.; Sharp, P. A. *T.I.B.S.* **1985**, *10*, 154-157.

(238) Wu, T.; Ogilvie, K. K.; Pon, R. T. *Nucl. Acids Res.* **1989**, *17*, 3501-3517.

(239) Kuijpers, W. H. A.; Huskens, J.; Koole, L. H.; van Boeckel, C. A. A. *Nucl. Acids Res.* **1990**, *18*, 5197-5205.

(240) Hilmoe, R. J. *J. Biol. Chem.* **1960**, *235*, 2117-2121.

(241) Wada, T.; Mochizuki, A.; Sato, Y.; Sekine, M. *Tetrahedron Lett.* **1998**, *39*, 7123-7126.

(242) Damha, M. J., personal communication.

(243) Kim, J.-W.; Kim, H.-C.; Kim, G.-M.; Yang, J.-M.; Boeke, J. D.; Nam, K. *Nucl. Acids Res.* **2000**, *28*, 3666-3673.

(244) Diener, T. O.; Owens, R. A.; Hammond, R. W. *Intervirol.* **1993**, *35*, 186-195.

(245) Ulanovsky, L.; Bodner, M.; Trifonov, E. N.; Choder, M. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 862-866.

(246) Egli, M.; Gessner, R. V.; Williams, L., D,; Quigley, G. J.; van der Marel, G. A.; van Boom, J. H.; Rich, A.; Frederick, C. A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 3235-3239.

(247) Prakash, G.; Kool, E. T. *J. Chem. Soc., Chem. Commun.* **1991**, 1161-1163.

(248) Prakash, G.; Kool, E. T. *J. Am. Chem. Soc.* **1992**, *114*, 3523-3527.

(249) Wang, S.; Kool, E. T. *Nucl. Acids Res.* **1994**, *22*, 2326-2333.

(250) Kool, E. T. *J. Am. Chem. Soc.* **1991**, *113*, 6265-6266.

(251) Rubin, E.; McKee, T. L.; Kool, E. T. *J. Am. Chem. Soc.* **1993**, *115*, 360-361.

(252) Rumney, S.; Kool, E. T. *Angew. Chem.* **1992**, *104*, 1686-1689.

(253) Fire, A.; Xu , S. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 4641-4645.

(254) Daubendiek, S. L.; Ryan, K.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, *117*, 7818-7819.

(255) Clusel, C.; Ugarte, E.; Enjolras, N.; Vasseur, M.; Blumenfeld, M. *Nucl. Acids Res.* **1993**, *21*, 3405-3411.

(256) Lim, C. S.; Jabrane-Ferrat, N.; Fontes, J. D.; Okamoto, H.; Garovoy, M. R.; Peterlin, B. M.; Hunt, C. A. *Nucl. Acids Res.* **1997**, *25*, 575-581.

(257) Hosoya, T.; Takeuchi, H.; Kanesaka, Y.; Yamakawa, H.; Miyano-Kurosaki, N.; Takai, K.; Yamamoto, N.; Takaku, H. *FEBS Lett.* **1999**, *461*, 136-40.

(258) Dolinnaya, N. G.; Blumenfeld, M.; Merenkova, I. N.; Oretskaya, T.; Krynetskaya, N. F.; Ivanoskaya, M. G.; Vasseur, M.; Shabarova, Z., A. *Nucl. Acids Res.* **1993**, *21*, 5403-5407.

(259) Nilsson, M.; Malmgren, H.; Samiotaki, M.; Kwiatkowski, M.; Chowdhary, B. P.; Landegren, U. *Science* **1994**, *265*, 2085-2088.

(260) Park, W.-S.; Miyano-Kurosaki, N.; Abe, T.; Takai, K.; Yamamoto, N.; Takaku, H. *Biochem. Biophys. Res. Comm.* **2000**, *270*, 953-960.

(261) Rowley, P. T.; Kosciolek, B. A.; Kool, E. T. *Mol. Med.* **1999**, *5*, 693-700.

(262) Kaufmann, G.; Klein, T.; Littauer, U. Z. *FEBS Lett.* **1974**, *46*, 271-275.

(263) Sugino, A.; Snopek, T. J.; Cozzarelli, N. R. *J. Biol. Chem.* **1977**, *252*, 1732-1738.

(264) Capobianco, M. L.; Carcuro, A.; Tondelli, L.; Garbesi, A.; Bonora , B. M. *Nucl. Acids Res.* **1990**, *18*, 2661-2669.

(265) De Napoli, L.; Galeone, A.; Mayol, L.; Messere, A.; Montesarchio, D.; Piccialli, G. *Bioorg. Med. Chem.* **1995**, *3*, 1325-1329.

(266) Alazzouzi, E.; Escaja, N.; Straub, M.; Marti, B.; Grandas, A.; Pedroso, E. *Nucleosides & Nucleotides* **1997**, *16*, 1513-1514.

(267) Alazzouzi, E.; Escaja, N.; Grandas, A.; Pedroso, E. *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 1506-1508.

(268) Frieden, M.; Grandas, A.; Pedroso, E. *Chem. Comm.* **1999**, 1593-1594.

(269) Ashley, G. W.; Kushlan, D. M. *Biochemistry* **1991**, *30*, 2927-2933.

(270) Kuznetsova, S. A.; Merenkova, I. N.; Kanevsky, I. E.; Shabarova, Z. A.; Blumenfeld, M. *Antisense & Nucl. Acid Drug Dev.* **1999**, *9*, 95-100.

(271) Shabarova, Z. A.; Fedorova, O. A.; Dolinnaya, N. G.; Gottikh, M. B. *Orig. Life & Evol. Bio.* **1997**, *27*, 555-566.

(272) Diegelman, A. M.; Kool, E. T. *Current Protocols in Nucleic Acid Chemistry* **2000**, *5.2.*

(273) Rubin, E.; Rumney IV, S.; Wang, S.; Kool, E. T. *Nucl. Acids Res.* **1995**, *23*, 3547-3553.

(274) Aymani, J.; Coll, M.; van der Marel, G.; van Boom, J. H.; Wang, A. H.-J.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 2526-2530.

(275) Maksimenko, A. V.; Gottikh, M. B.; Volkov, E. M.; Shabarova, Z. A. *Mol. Biol.* **1997**, *31*, 715-721.

(276) Li, T.; Liu, D.; Chen, J.; Lee, A. H. F.; Qi, J.; Chan, A. S. C. *J. Am. Chem. Soc.* **2001**, *123*, 12901-12902.

(277) Liu, D.; Chen, J.; Lee, A. H. F.; Chow, L. M. C.; Chan, A. S. C.; Li, T. *Angew. Chem. Int. Ed.* **2003**, *42*, 797-799.

(278) Rentzeperis, D.; Ho, J.; Markey, L. A. *Biochemistry* **1993**, *32*, 2564-2572.

(279) Wemmer, D. E.; Benight, A. S. *Nucl. Acids Res.* **1985**, *13*, 8611-8621.

(280) Amaratunga, M.; Snowden-Ifft, E.; Wemmer, D. E.; Benight, A. S. *Biopolymers* **1992**, *32*, 865-879.

(281) Olivera, B. M.; Scheffler, I. E.; Lehman, I. R. *J. Mol. Biol.* **1968**, *36*, 275-285.

(282) Erie, D. A.; Jones, R. A.; Olson, W. K.; Sinha, N. K.; Breslauer, K. J. *Biochemistry* **1989**, *28*, 268-273.

(283) Lim, C. S.; Hunt, C. A. *Nucleosides & Nucleotides* **1997**, *16*, 41-51.

(284) Haasnoot, C. A. G.; Hilbers, C. W.; Van der Marel, G. A.; Van Boom, J. H.; Singh, U. C.; Pattabiraman, N.; Kollman, P. A. *J. Biomol. Sruct. Dynam.* **1986**, *3*, 843-57.

(285) Bloomers, M. J. J.; Walters, J. A. L. I.; Haasnoot, C. A. G.; Aelen, J. M. A.; van der Marel, G. A.; van Boom, J. H.; Hilbers, C. W. *Biochemistry* **1989**, *28*, 7491-7498.

(286) Erie, D.; Sinha, N.; Olson, W.; Jones, R.; Breslauer, K. *Biochemistry* **1987**, *26*, 7150-7159.

(287) Santa Lucia, J. J. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1460-1465.

(288) Zuker, M.; Mathews, D. H.; Turner, D. H. In *RNA Biochemistry and Biotechnology*; Barciszewski, J., Clark, B. F. C., Eds.; NATO ASI Series, Kluwer Academic Publishers: Norwell, MA, 1999.

(289) Giannaris, P. A.; Damha, M. J. *Nucl. Acids Res.* **1993**, *21*, 4742-4749.

(290) Wasner, M.; Arion, D.; Borkow, G.; Noronha, A.; Uddin, A. H.; Parniak, M. A.; Damha, M. J. *Biochemistry* **1998**, *37*, 7478-7486.

(291) Kandimalla, E. R.; Manning, A.; Zhao, Q.; Shaw, D. R.; Byrn, R. A.; Sasisekharan, V.; Agrawal , S. *Nucl. Acids Res.* **1997**, *25*, 370-378.

(292) Lieber, E.; Takashi, E. *J. Org. Chem.* **1961**, *26*, 4472-4479.

(293) Wincott, F.; Di Renzo, A.; Shaffer, C.; Grimm, S.; Tracz, D.*et al. Nucl. Acids Res.* **1995**, *23*, 2677-2684.

(294) Sproat, B.; Colonna, F.; Mullah, B.; Tsou, C.; Andrus, A.; Hampel, A.; Vinayak, R. *Nucleosides & Nucleotides* **1995**, *14*, 255-273.

(295) Vargeese, C.; Carter, J.; Yegge, J.; Krivjansky, S.; Settle, A.; Kropp, E.; Peterson, K.; Pieken, W. *Nucl. Acids Res.* **1998**, *26*, 1046-1050.

(296) Horn, T.; Urdea, M. S. *Tetrahedron Lett.* **1986**, *27*, 4705-4708.

(297) Beld, A.; Claesen, C. A. A.; Roersma, E. S.; Schippers, W. J. M.; Keizer, L. M.; Tesser, G. I. *J. Royal Netherlands Chem. Soc.* **1984**, *103*, 196-202.

(298) Premraj, B. J.; Yahtindra, N. J. *J. Biomol. Sruct. Dynam.* **1998**, *16*, 313-328.

(299) Premraj, B. J.; Patel, P. K.; Kandimalla, E. R.; Agrawal, S.; Hosur, R. V.; Yathindra, N. *Biochem. Biophys. Res. Comm.* **2001**, *283*, 537-543.

(300) Shabarova, Z. A.; Dolinnaya, N. G.; Drutsa, V. L.; Melnikova, N. P.; Purmal , A. A. *Nucl. Acids Res.* **1981**, *9*, 5747-5761.

(301) Dolinnaya, N. G.; Gryaznova, O. I.; Sokolova, N. I.; Shabarova, Z. A. *Bioorg. Khim.* **1986**, *12*, 921-928.

(302) Shabarova, Z. A.; Veiko, V. P.; Dolinnaya, N. G.; Drusta, V. L.; Metelev, V. G.; Oretskaya, T. S.; Purmal, A. A. *Bioorg. Khim.* **1987**, *13*.

(303) Shabarova, Z. A. *Biochimie* **1988**, *70*, 1323-1334.

(304) Gilham, P. T. *J. Am. Chem. Soc.* **1962**, *84*, 687-688.

(305) Kanaya, E.; Yanagawa, H. *Biochemistry* **1986**, *25*, 7423-7430.

(306) Ferris, J. P.; Huang, C.-H.; Hagan Jr., W. J. *Nucleosides & Nucleotides* **1989**, *8*, 407-414.

(307) Orgel, L. E.; Lohrmann, R. *Acc. Chem. Res.* **1974**, *7*, 368.

(308) Ferris, J. P.; Yanagawa, H.; Dudgeon, P. A.; Hagan, W. J., Jr.; Mallare, T. E. *Orig. Life & Evol. Bio.* **1984**, *15*, 29-43.

(309) Sokolova, N. I.; Ashirbekova, D. T.; Dolinnaya, N. G.; Shabarova, Z. A. *FEBS Lett.* **1988**, *232*, 153-155.

(310) Dolinnaya, N. G.; Sokolova, N. I.; Ashirbekova, D. T.; Shabarova, Z. A. *Nucl. Acids Res.* **1991**, *19*, 3067-3072.

(311) Wolters, M.; Wittig , B. *Nucl. Acids Res.* **1989**, *17*, 5163-5172.

(312) Merenkova, I. P.; Dolinnaya, N. G.; Oretskaia, T. S.; Sokolova, N. I.; Shabarova, Z. A. *Bioorg. Khim.* **1992**, *18*, 85-91.

(313) Merenkova, I. N.; Oretskaya, T. S.; Sokolova, N. I.; Shabarova, Z. A. *Bioorg. Khim.* **1993**, *19*, 1205-1214.

(314) Damha, M. J.; Ogilvie, K. K. *Biochemistry* **1988**, *27*, 6403-6416.

(315) Sund, C.; Agback, P.; Koole, L. H.; Sandstrom, A.; Chattopadhyaya, J. *Tetrahedron* **1992**, *48*, 695-718.

(316) Markiewicz, W. T.; Wyrzykiewicz, T. K. *Nucl. Acids Res.* **1989**, *17*, 7149-7158.

(317) Kumar, P.; Gupta, K. C.; Rosch, R.; Seliger, H. *Chem. Lett.* **1997**, *12*, 1231-1232.

(318) Yousaf, S. I.; Carroll, A. R.; Clarke, B. E. *Gene* **1984**, *27*, 309-313.

(319) Sprecher, C. A.; Baase, W. A.; Johnson, W. C. J. *Biopolymers* **1979**, *18*, 1109-1019.

(320) Kool, E. T. *Acc. Chem. Res.* **1998**, *31*, 502-510.

(321) Kool, E. T. *Chem. Rev.* **1997**, *97*, 1473-1487.

(322) Kandimalla, E. R.; Manning, A.; Lathan, C.; Byrn, R. A.; Agrawal, S. *Nucl. Acids Res.* **1995**, *23*, 3578-3584.

(323) Kuznetsova, S. A.; Ivanovskaia, M. G.; Shabarova, Z. A. *Bioorg. Khim.* **1987**, *13*, 1139-41.

(324) Hall, K. B.; McLauglin, L. W. *Biochemistry* **1991**, *30*, 10606-10613.

(325) Mellema, J. R.; Haasnoot, C. A.; van der Marel, G. A.; Wille, G.; van Boeckel, C. A.; van Boom, J. H.; Altona , C. *Nucl. Acids Res.* **1983**, *11*, 5717-5738.

(326) Sambrook, J. In *Molecular Cloning: A Laboratory Manual*; Sambrook, J., Maniatis, T., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., 1989; Vol. 1.

(327) Uhlenbeck, O. C.; Gumport, R. I. In *The Enzymes*; Boyer, P. D., Ed.; Academic Press Inc.: New York, NY, 1982; Vol. 15B.

(328) Hinton, D. M.; Baez, J. A.; Gumport, R. I. *Biochemistry* **1978**, *d17*, 5091-5097.

(329) Moseman McCoy, M. I.; Gumport, R. I. *Biochemistry* **1980**, *19*, 635-642.

(330) Hinton, D.; Gumport, R. I. *Nucl. Acids Res.* **1979**, *7*, 453-464.

(331) Charlebois, A.-F.; Prior, T. K.; Hoadley, K. A.; Silverman, S. K. *J. Am. Chem. Soc.* **2003**, *125*, 5346-5350.

(332) Wang, Y.; Silverman, S. K. *J. Am. Chem. Soc.* **2003**, *125*, 6880-6881.

(333) Gottlieb, M. S. In *Morbidity and Mortality Weekly Report*, 1981; Vol. 30.

(334) Gottlieb, M. S.; Schroff, R.; Schanker, H. M. *N. Engl. J. Med.* **1981**, *305*, 1425-1431.

(335) Siegel, F. P.; Lopez, C.; Hammer, G. S.; Brown, A. E.; Kornfeld, S. J.*et al. N. Engl. J. Med.* **1981**, *305*, 1439-1444.

(336) Masur, H.; Michelis, M. A.; Greene, J. B.; Onorato, I.; Stouwe, R. A.*et al. N. Engl. J. Med.* **1981**, *305*, 1431-1438.

(337) Gottlieb, M. S. *N Engl J Med* **2001**, *344*, 1788-1791.

(338) Sepkowitz, K. A. *N Engl J Med* **2001**, *344*, 1764-1772.

(339) Chermann, J. C.; Barre-Sinoussi, F.; Dauguet, C.; Brun-Vezinet, F.; Rouzioux, C.; Rozenbaum, W.; Montagnier, L. *Antibio. Chemother.* **1983**, *32*, 48-53.

(340) Barre-Sinoussi, F.; Chermann, J. C.; Rey, F.; Nugeyre, M. T.; Chamaret, S.*et al. Science* **1983**, *220*, 868-871.

(341) Clavel, F.; Guetard, D.; Brun-Vezinet, F.; Chamaret, S.; Rey, M. A.*et al. Science* **1986**, *233*, 343-346.

(342) Clavel, F.; Guyader, M.; Guetard, D.; Salle, M.; Montagnier, L.; Alizon, M. *Nature* **1986**, *324*, 691-695.

(343) Whittle, H.; Morris, J.; Todd, J.; Corrah, T.; Sabally, S.; Bangali, J.; Ngom, P. T.; Rolfe, M.; Wilkins, A. *AIDS* **1994**, *8*, 1617-1620.

(344) Arts, E. J.; Wainberg, M. A. *Adv. Virus. Res.* **1998**, *46*, 97-163.

(345) Persaud, D.; Zhou, Y.; Siliciano, J. M.; Siliciano, R. F. *J. Virol.* **2003**, *77*, 1659-1665.

(346) Parniak, M. A.; Sluis-Cremer, N. *Adv. Pharmacol.* **2000**, *49*, 67-109.

(347) Hoggard, P. G.; Back, D. J. *Curr. Opin. Infect. Diseases* **2002**, *15*, 3-8.

(348) Gotte, M.; Wainberg, M. A. *Drug Resist. Updates* **2000**, *31*, 30-38.

(349) Hhu, M.; Wainberg, M. A. *J. Human Virol.* **2000**, *3*, 16-26.

(350) di Marzo Veronese, F.; Copeland, T. D.; DeVico, A. L.; Rahman, R.; Oroszlan, S.; Gallo, R. C.; Sarngadharan, M. G. *Science* **1986**, *231*, 1289-1291.

(351) Rodgers, D. W.; Gamblin, S. J.; Harris, B. A.; Ray, S.; Culp, J. S.; Hellmig, B.; Woolf, D. J.; Debouck, C.; Harrison, S. C. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1222-1226.

(352) Kohlstaedt, L. A.; Wang, J.; Friedman, J. M.; Rice, P. A.; Steitz, T. A. *Science* **1991**, *256*.

(353) Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.*et al. Nature Struc. Biol.* **1995**, *2*, 293-302.

(354) Jacobo-Molina, A.; Ding, J.; Nanni, R. G.; Clark Jr, A. D.; Lu, X.*et al. Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6320-6324.

(355) Huang, H.; Chopra, R.; Verdine, G. L.; Harrison, S. C. *Science* **1998**, *282*.

(356) Hansen, J.; Schulze, T.; Moelling, K. *J. Biol. Chem.* **1987**, *262*, 12393-12396.

(357) Davies II, J. F.; Hostomdka, Z.; Hostomsky, Z.; Jordan, S. R.; Mathhews, D. A. *Science* **1991**, *252*, 88-95.

(358) Cirino, N. M.; Cameron, C. E.; Smith, J. S.; Roth, M. J.; Benkovic, S. J.; Le Grice, S. F. J. *Biochemistry* **1995**, *34*, 9936-9943.

(359) Pari, K.; Mueller, G. A.; DeRose, E. F.; Kirby, T. W.; London, R. E. *Biochemistry* **2003**, *42*, 639-650.

(360) Davies, R. W. *Bioscience Reports* **1984**, *4*, 707-727.

(361) Smerdon, S.; Jager, J.; Wang, J.; Kohlstaedt, L.; Chirino, A.; Friedman, J.; Rice, P.; Steitz, T. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3911-3915.

(362) Majumdar, C.; Stein, C. A.; Cohen, J. S.; Broder, S.; Wilson, S. H. *Biochemistry* **1989**, *28*, 1340-1346.

(363) Dirani-Diab, R. E.; Sarih-Cottin, L.; Delord, B.; Dumon, B.; Moreau, S.; Toulme, J. J.; Fleury, H.; Litvak, S. *Antimicrob. Agents Chemother.* **1997**, *41*, 2141-2148.

(364) Tuerk, C.; MacDougal, S.; Gold, L. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 6988-6992.

(365) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505-510.

(366) Mizrahi, V.; Brooksbank, R. L.; Nkabinde, N. C. *J. Biol. Chem.* **1994**, *269*, 19245-19249.

(367) Tarrago-Litvak, L.; Andreola, M. L.; Fournier, M.; Nevinsky, G. A.; Parissi, V.; de Soultrait, V. R.; Litvak, S. *Current Pharmaceutical Design* **2002**, *8*, 595-614.

(368) Zhan, X.; Tan, C. K.; Scott, W. A.; Mian, M.; Downey, K. M.; So, A. G. *Biochemistry* **1994**, *33*, 1366-1372.

(369) Tan, C. K.; Civil , R.; Mian, A. M.; So, A. G.; Downey, K. M. *Biochemistry* **1991**, *30*, 4831-4835.

(370) Borkow, G.; Fletcher, R. S.; Barnard, J.; Arion, D.; Motakis, D.; Dmitrienko, G. I.; Parniak, M. A. *Biochemistry* **1997**, *36*, 3179-3185.

(371) Loya, S.; Tal, R.; Kashman, Y.; Hizi, A. *Antimicrob. Agents Chemother.* **1990**, *34*, 2009-12.

(372) Loya, S.; Hizi, A. *J. Biol. Chem.* **1993**, *268*, 9323.

(373) Shaw-Reid, C. A.; Munshi, V.; Graham, P.; Wolfe, A.; Witmer, M.*et al. J. Biol. Chem.* **2003**, *278*, 2777-2780.

(374) Gao, W. Y.; Han, F. S.; Storm, C.; Egan, W.; Cheng , Y. C. *Mol. Pharmacol.* **1192**, *41*, 223-229.

(375) Andreola, M. L.; Pileur, F.; Calmels, C.; Ventura, M.; Tarrago-Litvak, L.; Toulme, J. J.; Litvak, S. *Biochemistry* **2001**, *40*, 10087-10094.

(376) Chen, H.; Gold, L. *Biochemistry* **1994**, *33*, 8746-8756.

(377) Mangos, M. M.; Damha, M. J. *Curr. Topics Med. Chem.* **2002**, *2*, 1147-1171.

(378) Schreiber, S. L. *Science* **2000**, *287*, 1964-1969.

(379) Lima, W. F.; Crooke, S. T. *Biochemistry* **1997**, *36*, 390-398.

(380) Gamper, H. B.; Reed, M. W.; Cox, T.; Virosco, J. S.; Adams, A. D.; Gall, A. A.; Scholler, J. K.; Meyer, R. B. J. *Nucl. Acids Res.* **1993**, *21*, 145-150.

(381) Hannoush, R. N.; Damha, M. J. *J. Am. Chem. Soc.* **2001**, *123*, 12368-12374.

(382) Shaw, J.; Kent, K.; Bird, J.; Fishback, J.; Froehler, B. *Nucl. Acids. Res.* **1991**, *19*, 747-750.

(383) Tang, J.; Temsamani, J.; Agrawal, S. *Nucl. Acids. Res.* **1993**, *21*, 2729-2735.

(384) Kuwasaki, T.; Hosono, K.; Takai, K.; Ushijima, K.; Nakashima, H.; Saito, T.; Yamamoto, N.; Takaku, H. *Biochem. Biophys. Res. Comm.* **1996**, *228*.

(385) Cheong, C.; Varani, G.; Tinoco, I. J. *Nature* **1990**, *346*, 680-682.

(386) Varani, G.; Cheong, C.; Tinoco, I. J. *Biochemistry* **1991**, *30*, 3280-3289.

(387) Yamakawa, H.; Hosono, K.; Ishibashi, T.; Nakashima, H.; Inagawa, T.; Yamamoto, N.; Takai, K.; Takaku, H. *Nucleosides & Nucleotides* **1996**, *15*, 519-529.

(388) Yamakawa, H.; Abe, T.; Saito, T.; Kazuyuki, T.; Yamamoto, N.; Takaku, H. *Bioorg. Med. Chem.* **1998**, *6*, 1025-1032.

(389) Gopalakrishnan, V.; Peliska, J.; Benkovic, S. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 10763-10767.

(390) Gotte, M.; Fackler, S.; Hermann, T.; Perola, E.; Cellai, L.; Gross, H.; Le Grice, S.; Heumann, H. *EMBO J.* **1995**, *14*, 833-841.

(391) Schatz, O.; Cromme, F. V.; Gruninger-Leitch, F.; Le Grice, S. F. J. *FEBS Lett.* **1989**, *257*, 311-314.

(392) Hansen, J.; Schulze, T.; Mellert, W.; Moelling, K. *EMBO J.* **1988**, *7*, 239-243.

(393) Starnes, M.; Cheng, Y. *J. Biol. Chem.* **1989**, *264*, 7073-7077.

(394) Lamond, A. I.; Sproat, B. S. In *RNA Processing: A Practical Approach*; Higgins, S. J., Hames, B. D., Eds.; IRL Press: New York, 1994; Vol. I.

(395) Ryder, U.; Sproat, B. S.; Lamond, A. I. *Nucl. Acids Res.* **1990**, *18*, 7373-7379.

(396) Chaulk, S. G.; MacMillan, A. M. *Angew. Chem. Int. Ed.* **2001**, *40*, 2149-2152.

(397) Paterson, B. M.; Roberts, B. E.; Kuff, E. L. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 4370-4374.

(398) Zamecnik, P. C.; Stephenson, M. L. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 280-284.

(399) Uhlmann, E.; Peyman, A. *Chem. Rev.* **1990**, *90*, 543-584.

(400) Dominski, Z.; Kole, R. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8673-8677.

(401) Sierakowska, H.; Sambade, M. J.; Agrawal, S.; Kole, R. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 12840-12844.

(402) Friedman, K. J.; Kole, J.; Cohn, J. A.; Knowles, M. R.; Silverman, L. M.; Kole, R. *J. Biol. Chem.* **1999**, *274*, 36193-36199.

(403) Mann, C. J.; Honeyman, K.; Cheng, A. J.; Ly, T.; Lloyd, F.; Fletcher, S.; Morgan, J. E.; Partridge, T. A.; Wilton, S. D. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 42-47.

(404) Nilsen, T. W. *Cell* **1994**, *78*, 1-4.

(405) Reed, R. *Curr. Opin. Cell Biol.* **2000**, *12*, 340-345.

(406) Tuschl, T.; Sharp, P. A.; Bartel, D. P. *RNA* **2001**, *7*, 29-43.

(407) Jacquier, A.; Rodriguez, J. R.; Rosbash, M. *Cell* **1985**, *43*, 423-430.

(408) Berglund, A. J.; Chua, K.; Abovich, N.; Reed, R.; Rosbash, M. *Cell* **1997**, *89*, 781-787.

(409) Arning, S.; Gruter, P.; Bilbe, G.; Kramer, A. *RNA* **1996**, *2*, 794-810.
(410) Rutz, B.; Seraphin, B. *RNA* **1999**, *5*, 819-831.

(411) Berglund, J. A.; Rosbash, M.; Schultz, S. C. *RNA* **2001**, *7*, 682-691.

(412) MacMillan, A. M.; Query, C. C.; Allerson, C. R.; Chen, S.; Verdine, G. L.; Sharp, P. A. *Genes Dev.* **1994**, *8*, 3008.

(413) Chiara, M. D.; Gozani, O.; Bennett, M.; Champion-Arnaud, P.; Palandjian, L.; Reed, R. *Mol. Cell. Biol.* **1996**, *7*.

(414) Query, C. C.; Strobel, S. A.; Sharp, P. A. *Nucleic Acids Symposium Series* **1995**, *33*, 224-225.

(415) Lin, R.-J.; Newman, A. J.; Cheng, S.-C.; Abelson, J. *J. Biol. Chem.* **1985**, *260*, 14780-14792.

(416) Hernandez, N.; Keller, W. *Cell* **1983**, *35*, 89-99.

(417) Krainer, A. R.; Maniatis, T.; Ruskin, B.; Green, M. R. *Cell* **1984**, *36*, 993-1005.

(418) Schultz, M. C.; Choe, S. Y.; Reeder, R. H. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 1004-1008.

(419) Fouser, L. A.; Friesen, J. D. *Cell* **1986**, *45*, 81-93.

(420) Damha, Masad J.; Giannaris, P. A.; Marfey, Peter; Reid, L. S. *Tetrahedron Lett.* **1991**, *32*, 2573-2576.

(421) Damha, M. J.; Giannaris, P. A.; Marfey, P. *Biochemistry* **1994**, *33*, 7877-7885.

(422) Fontanel, M. L.; Bazin, H.; Teoule, R. *Anal. Biochem.* **1993**, *214*, 338-340.

(423) Frauendorf, C.; Hausch, F.; Rohl, I.; Lichte, A.; Vonhoff, S.; Klussmann, S. *Nucl. Acids. Res.* **2003**, *31*, e34.

(424) Ghetti, A.; Abelson, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 11461-11464.

(425) Moore, M. J.; Sharp, P. A. *Science* **1992**, *256*, 992-997.

(426) Barabino, S. M. L.; Sproat, B. S.; Lamond, A. I. *Nucl. Acids Res.* **1992**, *20*, 4457-4464.

(427) Rain, J.-C.; Legrain, P. *EMBO J.* **1997**, *16*, 1759-1771.

(428) Goffeau, A.; Barrell, B. G.; Bussey, H.; Davis, R. W.; Dujon, B.*et al. Science* **1996**, *274*, 546-567.

(429) Stevens, S. W.; Abelson, J. *Methods Enzymol.* **2002**, *351*, 200-220.

(430) Dignam, J. D.; Lebovitz, R. M.; Roeder, R. G. *Nucl. Acids Res.* **1983**, *11*, 1475-1489.

(431) Beigelman, L.; McSwiggen, J. A.; Draper, K. G.; Gonzalez, C.; Jensen, K.*et al. J. Biol. Chem.* **1995**, *270*, 25702-25708.

(432) Freier, S. M.; Altmann, K.-H. *Nucl. Acids Res.* **1997**, *25*, 4429-4443.

(433) Manoharan, M. *Biochim. Biophys. Acta* **1999**, *1489*, 117-130.

(434) Kramer, M. F.; Coen, D. M. In *Current Protocols in Molecular Biology*; John Wiley and Sons:; Vol. 15.

(435) Tsujimoto, Y.; Croce, C. C. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*.

(436) Raff, M. *Nature* **1998**, *396*, 119-122.

(437) Lewis, R. In *The Scientist*, 2003; Vol. 17.

(438) Bakhshi, A.; Jensen, J. P.; Goldman, P.; Wright, J. J.; McBride, O. W.; Epstein, A. L.; Korsmeyer, S. J. *Cell* **1985**, *41*, 899-906.

(439) Yang, E.; Korsmeyer, S. J. *Blood* **1996**, *88*, 386-401.

(440) Kelekar, A.; thompson, C. B. *Trends Cell Biol.* **1998**, *8*, 324-330.

(441) Minn, A. J.; Boise, L. H.; Thompson, C. B. *J. Biol. Chem.* **1996**, *271*, 6306-6312.

(442) Rochaizx, P.; Krajewski, S.; Reed, J. C.; Bonnet, F.; Voigt, J. J. *J. Pathol.* **1999**, *187*.

(443) Kole, R. *Acta. Biochim. Pol.* **1997**, *44*, 231-238.

(444) Sierakowska, H.; Gorman, L.; Kang, S.-H.; Kole, R. *Methods Enzymol.* **1999**, *313*, 506-521.

(445) Jansen, B.; Zangemeister-Wittke, U. *The Lancet* **2002**, *3*, 672-683.

(446) Opalinska, J. B.; Gewirtz, A. M. *Nature Rev.* **2002**, *1*, 503-514.

(447) Jansen, B.; Wacheck, V.; Heere-Ress, E. *The Lancet* **2000**, *356*, 1728-1733.

(448) Mercatante, D. R.; Mohler, J. L.; Kole, R. *J. Biol. Chem.* **2002**, *277*, 49374-49382.

(449) Villemaire, J.; Dion, I.; Abou Elela, S.; Chabot, B. *in preparation* **2003**.

(450) Chabot, B.; Blanchette, M.; Lapierre, I.; La Branche, H. *Mol. Cell. Biol.* **1997**, *17*, 1776-1786.

(451) Kole, R.; Sazani, P. *Curr. Opin. Molec. Ther.* **2001**, *3*, 229-234.

(452) Still, W. C.; Kahn, M.; Mitra, A. *J. Org. Chem.* **1978**, *43*, 2293--2295.

(453) Ellington, A.; Pollard, J. D. J. *Purification of oligonucleotides using denaturing polyacrylamide gel electrophoresis*; John Wiley & Sons: New York, 1988.

(454) Chen, Z.; Ruffner, D. E. *Biotechniques* **1996**, *21*, 820-822.

(455) Asara, J. M.; Allison, J. *Anal. Chem.* **1999**, *71*, 2866-2870.

(456) Distler, A. M.; Allison, J. *Anal. Chem.* **2001**, *73*, 5000-5003.

(457) Kibler-Herzog, L.; Zon, G.; Whittier, G.; Shaikh, M.; Wilson, W. D. *Anti-Cancer Drug Design* **1993**, *8*, 65-79.

(458) Fletcher, R.; Holleshak, S.; Nagy, G.; Arion, D.; Borkow, G.; Gu, Z.; Wainberg, M. A.; Parniak, M. A. *Protein Expression and Purification* **1996**, *7*, 27-32.